AD-A280 477

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE July 1993	3. REPORT TYPE AND DATES COVERED Final Report - June 1989-Sept 1992

4. TITLE AND SUBTITLE
Human Factors for Flight Deck Certification Personnel

5. FUNDING NUMBERS
FA3E2/A3093

6. AUTHOR(S)
Kim Cardosi and M. Stephen Huntley

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Research and Special Programs Administration
John A. Volpe National Transportation Systems Center
Cambridge, MA 02142-1093

8. PERFORMING ORGANIZATION REPORT NUMBER
DOT-VNTSC-FAA-93-4

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Federal Aviation Administration
Research and Development Service
800 Independence Ave., S.W.
Washington, DC 20591

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DOT/FAA/RD-93/5

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the National Technical Information Service, Springfield, VA 22161

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This document is a compilation of proceedings and lecture material on human performance capabilities that was presented to FAA flight deck certification personnel. A five-day series of lectures was developed to provide certification specialists with information on fundamental characteristics of the human operator that are relevant to flight deck operations. The series was designed to proceed from the presentation of basic information on human sensory capabilities, through human cognition, to the application of this knowledge to the design of controls and displays in the automated cockpit. The initial lectures were prepared and presented by published academic researchers. The later ones were presented by senior human factors practitioners employed by major American airframe manufacturers.

14. SUBJECT TERMS
Human Factors, Cockpit, Automation, Display Design, Human Performance, Human Engineering, Perception, Sensation, Attention, Workload, Evaluation

15. NUMBER OF PAGES
416

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

U.S. Department
of Transportation

**Research and
Special Programs
Administration**

John A. Volpe
National Transportation
Systems Center

Kendall Square
Cambridge Massachusetts 02142

May 1994

**Human Factors for Flight Deck Certification Personnel
Final Report - July 1993**

ERRATA

Due to an oversight on the part of the printer, certain last-minute changes to this Final Report were not added. As a result, it is necessary for us to enclose a revised copy of page 44. Please accept our apologies. In the future, we intend to print a new edition of ***Human Factors for Flight Deck Certification Personnel*** that contains these additional changes.

that color discriminations that depend on S cones will be impaired if the image is sufficiently small to fall only on the center of the fovea. This is illustrated by Figure 3.3. When viewed close, so that the visual angle of each circle subtends several degrees, it is easy for an individual with normal color vision to discriminate the yellow vs. white and red vs. green. Viewed from a distance of several feet, however, the yellow and white will be indiscriminable. This is called ***small-field tritanopia***, because tritanopes are individuals who completely lack S cones. A tritanope would not be able to discriminate the yellow from the white in Figure 3.3 regardless of their sizes. With certain small fields, even normal individuals behave like tritanopes. Notice that even from a distance, the red-green pair is still discriminable because S cones are not necessary for this discrimination. Thus, the small-field effect is limited to discriminations that depend on S cones. (Note: Due to technical difficulties in reproducing colors, individuals with normal color vision may still be able to discriminate the yellow and white semicircles at a distance.)

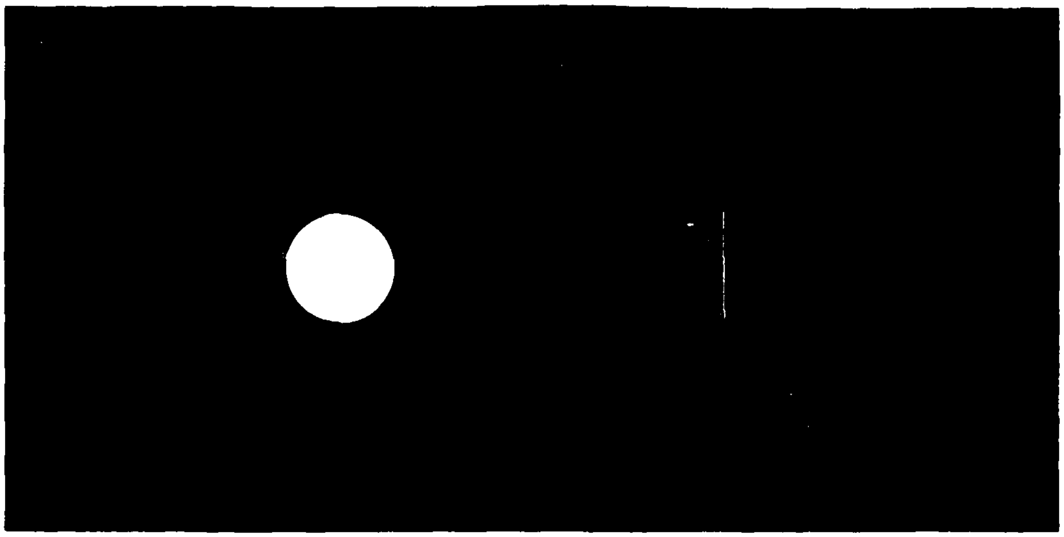


Figure 3.3. Colors (yellow and white) not discriminable at a distance due to small field tritanopia.

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in.) = 2.5 centimeters (cm)
1 foot (ft) = 30 centimeters (cm)
1 yard (yd) = 0.9 meter (m)
1 mile (mi) = 1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2) = 6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2) = 0.09 square meter (m^2)
1 square yard (sq yd, yd^2) = 0.8 square meter (m^2)
1 square mile (sq mi, mi^2) = 2.6 square kilometers (km^2)
1 acre = 0.4 hectares (he) = 4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz) = 28 grams (gr)
1 pound (lb) = .45 kilogram (kg)
1 short ton = 2,000 pounds (lb) = 0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp) = 5 milliliters (ml)
1 tablespoon (tbsp) = 15 milliliters (ml)
1 fluid ounce (fl oz) = 30 milliliters (ml)
1 cup (c) = 0.24 liter (l)
1 pint (pt) = 0.47 liter (l)
1 quart (qt) = 0.96 liter (l)
1 gallon (gal) = 3.8 liters (l)
1 cubic foot (cu ft, ft^3) = 0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3) = 0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

[(x - 32) (5/9)] °F = y°C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm) = 0.04 inch (in)
1 centimeter (cm) = 0.4 inch (in)
1 meter (m) = 3.3 feet (ft)
1 meter (m) = 1.1 yards (yd)
1 kilometer (km) = 0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2) = 0.16 square inch (sq in, in^2)
1 square meter (m^2) = 1.2 square yards (sq yd, yd^2)
1 square kilometer (kn^2) = 0.4 square mile (sq mi, mi^2)
1 hectare (he) = 10,000 square meters (m^2) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gr) = 0.036 ounce (oz)
1 kilogram (kg) = 2.2 pounds (lb)
1 tonne (t) = 1,000 kilograms (kg) = 1.1 short tons

VOLUME (APPROXIMATE)

1 milliliter (ml) = 0.03 fluid ounce (fl oz)
1 liter (l) = 2.1 pints (pt)
1 liter (l) = 1.06 quarts (qt)
1 liter (l) = 0.26 gallon (gal)
1 cubic meter (m^3) = 36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3) = 1.3 cubic yards (cu yd, yd^3)

TEMPERATURE (EXACT)

[(9/5) y + 32] °C = x °F

QUICK INCH-CENTIMETER LENGTH CONVERSION

INCHES	0	1	2	3	4	5	6	7	8	9	10

CENTIMETERS: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 25.40

QUICK FAHRENHEIT-CELSIUS TEMPERATURE CONVERSION

°F	-40°	-22°	-4°	14°	32°	50°	68°	86°	104°	122°	140°	158°	176°	194°	212°
°C	-40°	-30°	-20°	-10°	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°

For more exact and or other conversion factors, see NBS Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50. SD Catalog No. C13 10286.

Preface

Flight test pilots who perform aircraft certification and evaluation functions for the FAA are frequently required to make important decisions regarding the human factors aspects of cockpit design. Such decisions require a thorough understanding of the operational conditions under which cockpit systems are used as well as the performance limits and capabilities of the flight crews who will use these systems. In the past, the limits of control and display technology and the test pilot familiarity with the knobs and dials of traditional aircraft have provided useful references from which to judge the safety and utility of cockpit displays and controls. Today, however, with the advent of the automated cockpit, and the almost limitless information configurations possible with CRT and LCD displays, evaluators are being asked to go far beyond their personal experience to make certification judgments.

A survey of human factors handbooks, advisory circulars and even formal human factors courses revealed little material on human performance that was formatted in a fashion that would provide useful guidelines to certification personnel for human factors evaluations in the cockpit. Most sources of human factors information are of limited use in evaluating advanced technology cockpits because they are out of date and do not consider the operational and cockpit context within which the newly designed controls and displays are to be used.

It will be some time before the human factors issues concerning interacting with electronic cockpits are well defined and there is sufficient information and understanding available to support the development of useful handbooks. In lieu of such guidance, a series of one-week seminars on human factors issues relevant to cockpit display design was conducted for approximately 120 FAA certification personnel. The lectures were given by researchers and practitioners working in the field. The lectures included material on the special abilities and limitations of the human perceptual and cognitive system, concepts in display design, testing and evaluation, and lessons learned from the designers of advanced cockpit display systems. The contents of this document were developed from the proceedings of the seminars.

I wish to thank a number of my friends and colleagues for their important contribution to this document. I am deeply indebted to Dr. Kim Cardosi, Dr. Peter D. Eimas, Mr. Delmar M. Fadden, Dr. Richard F. Gabriel, Dr. Christopher D. Wickens, and Dr. John Werner, the principal

authors of the material in this book. Each of them is a respected and highly productive professional in his or her own field and has contributed rare and valuable time to this activity. Clearly there would be no document without their contributions.

I am particularly grateful to Dr. Kim Cardosi, the project manager, for her editing and able administration of much of the work that culminated in this report. This work included the management of the four seminars as well as the organization of the resulting proceedings into the textbook format of the current document.

Special thanks are due to Mr. Paul McNeil, Mr. Arthur H. Rubin, and Mr. Jim Green of EG&G Dynatrend Corp. for their many hours and tireless efforts in assembling and publishing the manuscripts included herein, and to Ms. Rowena Morrison of Battelle for her insightful and thoughtful support in editing particularly troublesome sections of this work.

The four seminars and the publishing of the resulting report were generously funded through the Federal Aviation Administration's Flight Deck Human Factors Research Program managed by Mr. William F. White.

M.S. Huntley, Ph.D.
Manager, Cockpit Human Factors Program

CONTENTS

LIST OF FIGURES

List of Tables

ACKNOWLEDGEMENT OF PERMISSION TO REPRINT

FIGURES

Figure 1.2 From Fletcher, H. ***Speech and Hearing in Communication***. (2nd edit.) (Copyright • 1953) New York: Van Nostrand Co. Reproduced by permission of the publisher.

Figure 1.3 From Fletcher, H. & Munson, W.A. Loudness, its definition, measurement and calculation. (Copyright • 1933) ***Journal of the Acoustical Society of America***, 5, 82-108. Reproduced by permission of the publisher.

Figure 1.4 From Werner, J.S. & Schlesinger, K. ***Psychology: Science of Mind, Brain and Behavior***. (Copyright • 1991) New York: McGraw-Hill. Reproduced by permission of McGraw-Hill.

Figure 2.1 From figures in ***Sensation and Perception***, Second Edition by Stanley Coren, Clare Porac, and Lawrence M. Ward, copyright • 1984 by Harcourt Brace & Company. Reproduced by permission of the publisher.

Figure 2.2 From Wyszecki, G. & Stiles, W.S. ***Color Science: Concepts and Methods, Quantitative Data and Formulae***, (2nd ed.) (Copyright • 1982) New York: John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

Figure 2.3 From Walraven, J. The colours are not on the display: a survey of non-veridical perceptions that may turn up on a colour display. This figure and article were first published in ***Displays***, Vol 6, January 1985, pp 35-42 and the figure is reproduced here with the permission of Butterworth-Heinemann, Oxford, UK.

Figure 2.4 From figures from ***Visual Perception*** by Tom N. Cornsweet, copyright • 1970 by Harcourt Brace & Company. Reproduced by permission of the publisher.

Figure 2.5 From figures in ***Sensation and Perception***, Second Edition, by Stanley Coren, Clare Porac, and Lawrence M. Ward, copyright • 1984 by Harcourt Brace & Company. Reproduced by permission of the publisher.

Figure 2.6 From Helps, E.P.W. Physiological effects of aging (copyright • 1973). ***Proceedings of the Royal Society of Medicine***, 66, 815-818. Reproduced by permission of the Journal of the Royal Society of Medicine.

Figure 2.8	From Werner, J.S., Peterzell, D. & Scheetz, A.J. Light, vision, and aging -- A brief review. (Copyright • 1990) ***Optometry and Vision Science***, 67 214-229. Reproduced by permission of Williams & Wilkins, American Academy of Optometry.
Figure 2.9	From Wyszecki, G. & Stiles, W.S. ***Color Science: Concepts and Methods, Quantitative Data and Formulae.*** (2nd ed.) (Copyright • 1982) New York: John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.
Figure 2.10	From Osterberg, G. Topography of the layer of rods and cones in the human retina. ***Acta Ophthalmologica***, Supplement, (copyright • 1935). Reproduced by permission of Scriptor Publisher ApS, Copenhagen, Denmark.
Figure 2.12	From Judd, D.B. Basic correlates of the visual stimulus. In S.S. Stevens (Ed.) ***Handbook of Experimental Psychology***, (copyright • 1951) New York: John Wiley & Sons, Inc., 811-867. Reproduced by permission of John Wiley & Sons, Inc.
Figure 2.13	From Graham, C.H. Some fundamental data. In C.H. Graham (Ed.), ***Vision and Visual Perception***. (Copyright • 1965a) New York: John Wiley & Sons, Inc., 68-80. Reproduced by permission of John Wiley & Sons, Inc.
Figure 2.14	From Hecht, S. Vision: II. The nature of the photoreceptor process. In C. Murchison (Ed.), ***A Handbook of General Experimental Psychology***. (Copyright • 1934) Worcester, MA: Clark University Press, 704-828. Reproduced by permission of the publisher.
Figure 2.16	From Sekuler, R. & Blake, R. ***Perception***. (Copyright • 1985) New York: McGraw-Hill, Inc. Reproduced by permission of McGraw-Hill.
Figure 2.17	From Yarbus, A.L. ***Eye Movements and Vision***. (Copyright • 1967) New York: Plenum Publishing Corp. Reproduced by permission of the publisher.
Figure 2.18	From Ditchburn, R.W. Eye movements in relation to retinal action. (Copyright • 1955) ***Optica Acta***, 1, 171-176. Reproduced by permission of Taylor & Francis Ltd., London, U.K.
Figure 2.19	From Polyak, S. ***The Retina***. (Copyright • 1941) Chicago: University of Chicago Press. Reproduced by permission of the publisher. All rights reserved.

Figure 2.20 From Hecht, S. & Smith, E.L. Intermittent stimulation by light: VI. Area and the relation between critical frequency and intensity. Reproduction from the *Journal of General Physiology*, 1936, 19, 979-989. Reproduced by copyright permission of the Rockefeller University Press.

Figure 2.21 From Hecht, S. & Verrijp, C.D. Intermittent stimulation by light. III. The relation between intensity and critical fusion frequency for different retinal locations. Reproduction from the *Journal of General Physiology*, 1933, 17, 251-265. Reproduced by copyright permission of the Rockefeller University Press.

Figure 2.22 From Weale, R.A. *A Biography of the Eye*. (Copyright © 1982) London: H.K. Lewis & Co. Reproduced by permission of Chapman & Hall, London, U.K.

Figure 2.23 From deLange, H. Research into the nature of the human fovea-cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. (Copyright © 1958) Washington, D.C.: *Journal of the Optical Society of America*, 48, 777-784. Reproduced by permission of the publisher.

Figure 2.25 From Goldstein, E.B. *Sensation and Perception* 2/E, (copyright © 1984) by Wadsworth, Inc. Reproduced by permission of the publisher.

Figure 3.1 From Bowmaker, J.K., Dartnall, H.J.A. & Mollon, J.D. Microspectrophotometric demonstration of four classes of photoreceptors in an old world primate Macaca fascicularis. (Copyright © 1980) *Journal of Physiology*, 298, 131-143. Reproduced by permission of the Physiological Society, Oxford, U.K.

Figure 3.2 From Marc, R. E. & Sperling, H. G. Chromatic organization of primate cones. (Copyright © 1977 by AAAS) *Science*, 196, 454-456. Reproduced by permission of the AAAS, Washington, D.C.

Figure 3.4 From Werner, J.S. & Steele, V.G. Sensitivity of human foveal color mechanisms throughout the life span. (Copyright © 1988) Washington, D.C.: *Journal of the Optical Society of America*, A, 5, 2122-2130. Reproduced by permission of the publisher.

Figure 3.6 From the Farnsworth Dichotomous Test of Color Blindness, Panel D-15. (Copyright © 1947) The Psychological Corporation. Reproduced by permission of The Psychological Corporation. All rights reserved.

Figure 3.9	From Werner, J.S. & Wooten, B.R. Opponent-chromatic mechanisms: Relation to photopigments and hue naming. (Copyright © 1979) Washington, D.C.: ***Journal of the Optical Society of America***, 69, 422-434. Reproduced by permission of the publisher.
Figure 3.10	From Hurvich, L.M. ***Color Vision***. (Copyright © 1981) Sunderland, MA: Sinauer Associates. Reproduced by permission of Leo Hurvich.
Figure 3.11	From Volbrecht, V.J., Aposhyan, H.M. & Werner, J.S. Perception of electronic display colours as a function of retinal illuminance. (Copyright © 1988) ***Displays***, 9, 56-64. Reproduced by permission of Butterworth-Heinemann Ltd., London, U.K.
Figure 3.12	From Hurvich, L.M. ***Color Vision***. (Copyright © 1981) Sunderland, MA: Sinauer Associates. Reproduced by permission of Leo Hurvich.
Figure 3.13	From Wright, W.D. & Pitt, F.H.G. Hue discrimination in normal colour-vision. (Copyright © 1934) ***Proceedings of the Physical Society***, 46, 459-473. Reproduced by permission of the IOP, Bristol, U.K.
Figure 3.14	From Hurvich, L.M. ***Color Vision***. (Copyright © 1981) Sunderland, MA: Sinauer Associates. Reproduced by permission of Leo Hurvich.
Figure 3.15	From Albers, J. ***Interaction of Color***. (Copyright © 1975) New Haven, CT: Yale University Press. Reproduced by permission of the publisher.
Figure 3.16	From Evans, R.M. ***An Introduction to Color***. (Copyright © 1948) New York: John Wiley & Sons, Inc. Reproduced by courtesy of Pauline Evans.
Figure 3.17	From Werner, J.S. & Walraven, J. Effect of chromatic adaptation on the achromatic locus: The role of contrast, luminance and background color. (Copyright © 1982) ***Vision Research***, 22, 929-943. Reproduced by permission of Pergamon Press, plc., Elmsford, NY.
Figure 3.18	From Viveash, J.P. & Laycock, J. Computation of the resultant chromaticity coordinates and luminance of combined and filtered sources in display design. This figure and article were first published in ***Displays***, Vol 4, January 1983, pp 17-23. The figure was reproduced here with the permission of Butterworth-Heinemann, Oxford, UK.

Figure 3.19 From Wyszecki, G. & Stiles, W.S. (2nd ed.) ***Color Science: Concepts and Methods, Quantitative Data and Formulae.*** (Copyright © 1982) New York: John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

Figure 3.20 From Wyszecki, G. & Stiles, W.S. ***Color Science: Concepts and Methods, Quantitative Data and Formulae.*** (2nd ed.) (Copyright © 1982) New York: John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

Figure 3.21 From Volbrecht, V.J., Aposhyan, H.M. & Werner, J.S. Perception of electronic display colours as a function of retinal illuminance. (Copyright © 1988) ***Displays***, 9, 56-64. Reproduced by permission of Butterworth-Heinemann Ltd., London, U.K.

Figure 3.22 From Wyszecki, G. & Stiles, W.S. ***Color Science: Concepts and Methods, Quantitative Data and Formulae.*** (2nd ed.) (Copyright © 1982) New York: John Wiley & Sons, Inc. Reproduced by permission of John Wiley & Sons, Inc.

Figure 4.1 From figures in ***Visual Perception*** by Tom N. Cornsweet, copyright © 1970 by Harcourt Brace & Company. Reproduced by permission of the publisher.

Figure 4.2 From figures in ***Visual Perception*** by Tom N. Cornsweet, copyright © 1970 by Harcourt Brace & Company. Reproduced by permission of the publisher.

Figure 4.3 From figures in ***Visual Perception*** by Tom N. Cornsweet, copyright © 1970 by Harcourt Brace & Company. Reproduced by permission of the publisher.

Figure 4.4 From Campbell, F.W. & Robson, J.G. Application of Fourier analysis to the visibility of gratings. (Copyright © 1968) ***Journal of Physiology***, 197, 551-566. Reproduced by permission of the Physiological Society, Oxford, U.K.

Figure 4.5 From Blakemore, C.B. & Sutton, P. Size adaptation: A new aftereffect. (Copyright © 1969 by AAAS) ***Science***, 166, 245-247. Reproduced by permission of the AAAS, Washington, D.C.

Figure 4.6 From DeValois, R.L., Morgan, H. & Snodderly, D.M. Psychophysical studies of monkey vision--III. Spatial luminance contrast sensitivity tests of macaque and human observers. (Copyright © 1974) ***Vision Research***, 14, 75-81. Reproduced by permission of Pergamon Press, plc., Elmsford, NY.

Figure 4.7	From Rovamo, J., Virsu, V. & Näsänen, R. Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. (Copyright © 1978) ***Nature***, 271, 54-56. Reproduced by permission of Macmillan Magazines, London, U.K.
Figure 4.8	From Owsley, C., Sekuler, R. & Siemsen, D. Contrast sensitivity throughout adulthood. (Copyright © 1983) ***Vision Research***, 23, 689-699. Reproduced by permission of Pergamon Press, plc., Elmsford, NY.
Figure 4.9	From DeValois, R.L. & DeValois, K.K. ***Spatial Vision.*** (Copyright © 1988) New York: Oxford University Press. Reproduced by permission of the publisher.
Figure 4.10	From DeValois, R.L. & DeValois, K.K. ***Spatial Vision.*** (Copyright © 1988) New York: Oxford University Press. Reproduced by permission of the publisher.
Figure 4.11	From DeValois, R.L. & DeValois, K.K. ***Spatial Vision.*** (Copyright © 1988) New York: Oxford University Press. Reproduced by permission of the publisher.
Figure 4.14	From Sekuler, R. & Blake, R. ***Perception.*** (Copyright © 1985) New York: McGraw-Hill, Inc. Reproduced by permission of McGraw-Hill.
Figure 4.15	From Gibson, J.J. ***The Senses Considered as Perceptual Systems.*** Copyright © 1966. Reproduced by Houghton Mifflin Company. Used with permission.
Figure 4.16	From figures from ***Sensation and Perception,*** Second Edition by Stanley Coren, Clare Porac, and Lawrence M. Ward, copyright © 1984 by Harcourt Brace & Company. Reproduced by permission of the publisher.
Figure 4.17	From Matlin, M.W. ***Sensation and Perception.*** (2nd edit.) (copyright © 1983) Boston: Allyn and Bacon. Reproduced by permission of the publisher.
Figure 4.18	From Werner, J.S. & Schlesinger, K. ***Psychology: Science of Mind, Brain and Behavior.*** (Copyright © 1991) New York: McGraw-Hill. Reproduced by permission of McGraw-Hill.
Figure 4.19	From Julesz, B. ***Foundations of Cyclopean Perception.*** (Copyright © 1971) Chicago: University of Chicago Press. Reproduced by permission of the publisher. All rights reserved.

Figure 6.1 From Wickens, C.D. ***Engineering Psychology and Human Performance*** (2nd edit.). (Copyright © 1992) New York: HarperCollins. Reproduced by permission of HarperCollins, College Publishers.

Figure 6.2 From Wickens, C.D. ***Engineering Psychology and Human Performance*** (2nd edit.). (Copyright © 1992) New York: HarperCollins. Reproduced by permission of HarperCollins, College Publishers.

Figure 6.3 From Wickens, C.D. ***Engineering Psychology and Human Performance*** (2nd edit.). (Copyright © 1992) New York: HarperCollins. Reproduced by permission of HarperCollins, College Publishers.

Figure 6.4 From Desmond, J. ***Improvements in aircraft safety and operational dependability from a projected flight path guidance display,*** reproduced with permission from SAE Paper No. 861732, copyright © 1986, Society of Automotive Engineers, Inc.

Figure 6.5 From Desmond, J. ***Improvements in aircraft safety and operational dependability from a projected flight path guidance display,*** reproduced with permission from SAE Paper No. 861732, copyright © 1986, Society of Automotive Engineers, Inc.

Figure 7.1 From Wickens, C.D. ***Engineering Psychology and Human Performance*** (2nd edit.). (Copyright © 1992) New York: HarperCollins. Reproduced by permission of HarperCollins, College Publishers.

Figure 7.2 From Wickens, C.D. & Flach, J. Human Information Processing, in E. Weiner & D. Nagel (Eds.), ***Human Factors in Aviation***, 111-155, (copyright © 1988) Orlando, FL: Reproduced by permission of Academic Press.

Figure 7.3 From Wright, P. Presenting technical information: a survey of research findings. (Copyright © 1977) ***Instructional Science***, 6, 93-134. Reproduced by permission of Kluwer Academic Publishers, Dordrecht, Netherlands.

Figure 7.7 From Wickens, C.D. ***Engineering Psychology and Human Performance*** (2nd edit.). (Copyright © 1992) New York: HarperCollins. Reproduced by permission of HarperCollins, College Publishers.

Figure 7.8 From Hawkins, F.H. *Human Factors in Flight*, (copyright © 1987) Brookfield, VT: Gower Technical Press. Reproduced by permission of Ashgate Publishing Co., Brookfield, VT

Figure 7.9 From Braune, R.J. *The common/same type rating: human factors and other issues.* Reproduced with permission from SAE Paper No. 892229, copyright © 1989, Society of Automotive Engineers, Inc.

Figure 8.1 From Wickens, C.D. *Engineering Psychology and Human Performance* (2nd edit.). (Copyright © 1992) New York: HarperCollins. Reproduced by permission of HarperCollins, College Publishers.

Figure 8.3 From Parks, D.L. & Boucek, G.P., Jr. Workload prediction, diagnosis and continuing challenges. In G.R. McMillan et al. (Eds.), *Applications of Human Performance Models to System Design*, (copyright © 1989) New York: Plenum Publishing Corp. 47-63. Reproduced by permission of the publisher.

Figure 8.4 From Groce, J.L. & Boucek, G.P., Jr. *Air transport crew tasking in an ATC data link environment.* Reproduced with permission from SAE Paper No. 871764, copyright © 1987, Society of Automotive Engineers, Inc.

Figure 8.5 From Wickens, C.D. *Engineering Psychology and Human Performance* (2nd edit.). (Copyright © 1992) New York: HarperCollins. Reproduced by permission of HarperCollins, College Publishers.

Figure 8.7a From Roscoe, A.H. The original version of this material was first published by the Advisory Group for Aerospace Research and Development, North Atlantic Treaty Organisation (AGARD/NATO) in AGARDograph AG-282 "The practical assessment of pilot workload" in 1987.

Figure 8.7b From Cooper, G.E. & Harper, R.P., Jr. The original version of this material was first published by the Advisory Group for Aerospace Research and Development, North Atlantic Treaty Organisation (AGARD/NATO) in AGARDograph AG-567 "The use of pilot rating in the evaluation of aircraft handling qualities" in 1969.

Figure 8.8 From Wilson, G.F., Skelly, J., & Purvis, G. The original version of this material was first published by the Advisory Group for Aerospace Research and Development, North Atlantic Treaty Organisation (AGARD/NATO) in AGARDograph CP-458 "Reactions to emergency situations in actual and simulated flight. In Human Behavior in High Stress Situations in Aerospace Operations" in 1989.

Figure 8.9	From Wilson, G.F., Skelly, J., & Purvis, G. The original version of this material was first published by the Advisory Group for Aerospace Research and Development, North Atlantic Treaty Organisation (AGARD/NATO) in AGARDograph CP-458 "Reactions to emergency situations in actual and simulated flight. In Human Behavior in High Stress Situations in Aerospace Operations" in 1989.
Figure 8.11	From Czeisler, C.A., Weitzman, E.D., Moore-Ede, M.C., Zimmerman, J.C., & Knauer, R.S. Human sleep: its duration and organization depend on its circadian phase. (Copyright © 1980) ***Science***, 210, 1264-1267. Reproduced by permission of AAAS.
Figure 8.12	From Richardson et al. Circadian Variation in Sleep Tendency in Elderly and Young Subjects. (Copyright © 1982) ***Sleep***, 5 (suppl. 2), A.P.S.S., 82. Reproduced by permission of the publisher.
Figure 8.13	From Klein, K.G., Wegmann, H.M. & Hunt, B.I. Desynchronization of body temperature and performance circadian rhythms as a result of outgoing and homegoing transmeridian flights. (Copyright © 1972) ***Aerospace Medicine***, 43, 119-132. Reproduced by permission of Aerospace Medical Association, Alexandria, VA.
Figure 8.15	From Wegman, H.M., Gundel, A., Naumann, M., Samel, A., Schwartz, E. & Vejvoda, M. Sleep, sleepiness, and circadian rhythmicity in aircrews operating on transatlantic routes. (Copyright © 1986) ***Aviation, Space, and Environmental Medicine***, 57, (12, suppl.), B53-B64. Reproduced by permission of Aerospace Medical Association, Alexandria, VA.
Figure 9.2	From Nagel, D.C. Aviation safety: needs for human factors research. (Copyright © 1987) Presentation to Air Transport Association (ATA) Airlines Operations Forum, 19-21 October, Annapolis, Maryland. Reproduced by permission of the ATA.
Figure 9.4	From a document of the Boeing Commercial Airplane Group, Seattle, Washington, Reproduced by permission of Boeing Company.
Figure 10.1	From Society of Automotive Engineers, Human interface design methodology for integrated display symbology. (Copyright © 1990) from SAE document ARP 4155. Reproduced by permission of the Society of Automotive Engineers, Inc.
Figures 11.1 - 11.11	From a document of the Boeing Commercial Airplane Group, Seattle, Washington. Reproduced by permission of Boeing Company.

TABLES

Table 8.2 From Klein, K.G., Wegmann, H.M. & Hunt, B.I. Desynchronization of body temperature and performance circadian rhythms as a result of outgoing and homegoing transmeridian flights. (Copyright © 1972) ***Aerospace Medicine***, 43, 119-132. Reprinted by permission of Aerospace Medical Association, Alexandria, VA.

Table 9.3 From Barnett, A. & Higgins. M.K. Airline safety: the last decade. ***Management Science***, 35, January 1989, 1-21, (copyright © 1989). Reprinted by permission of The Institute of Management Sciences, 290 Westminster St., Providence, Rhode Island 02903, USA.

Table 9.7 From Bainbridge, L. Ironies of automation. In ***New Technology and Human Error***. Ed. M.J. Rasmussen, K. Duncan, and J. Leplat. (Copyright © 1987) West Sussex, U.K.: John Wiley & Sons Ltd. Reprinted by permission of John Wiley & Sons Ltd.

Table 9.8 From Meister, D. A cognitive theory of design and requirements for a behavioral design aid in system design. In ***Behavioral Perspectives for Designers, Tools and Organizations***. Ed. W.B. Rouse and K.R. Boff. (Copyright © 1987) New York: North Holland Publications. Reprinted by permission of W.B. Rouse.

Table 9.9 From a document of the Boeing Commercial Airplane Group, Seattle, Washington. Reprinted by permission of Boeing Company.

Executive Summary

A series of one-week seminars was developed to provide FAA certification specialists with information on fundamental characteristics of the human operator relevant to cockpit operations with examples of applications of this information to aviation problems. The series was designed to proceed from the development of basic information on human sensory capabilities, through human cognition, to the application of this knowledge to the design of controls and displays in the automated cockpit.

The earlier lectures were prepared and presented by published academic researchers, the later ones by human factors practitioners employed by the major airframe manufacturers in the United States.

The lecture series was presented on four separate occasions and was attended by approximately 120 FAA flight test and evaluation group pilots and engineers. This text is a compilation of the lecture material presented to these professionals during the four occasions.

Chapter 1

Auditory Perception

by John S. Werner, Ph.D., University of Colorado at Boulder

Hearing, like vision, provides information about objects and events at a distance. There are some important practical differences between hearing and vision. For example, the stimulus for vision, light, cannot travel through solid objects, but many sounds can. Unlike vision, hearing is not entirely dependent on the direction of the head. This makes auditory information particularly useful as a warning system. A pilot can process an auditory warning regardless of the direction of gaze, and while processing other critical information through the visual channel. **Auditory information is also less degraded than visual signals by turbulence** during flight, making auditory warnings an appropriate replacement for some visual display warnings (Stokes & Wickens, 1988). No doubt these considerations formed the basis of FAA voluntary guidelines on the use of aural signals as part of aircraft alerting systems (RD-81/38,II, page 89).

Physical Properties of Sound

Let us start by exploring what happens in the physical world to generate sound. When you pluck the string of a guitar, it vibrates back and forth compressing a small surrounding region of air. When the vibrating string moves away, it pushes air in the opposite direction, creating a region of decompression. As the string vibrates back and forth, it creates momentary increases and decreases in air pressure, or sound waves. These alternating increases and decreases travel through the air at a speed of approximately 740 miles per hour (Mach I, the speed of sound). Eventually they arrive at our ear, where the tympanic membrane, our eardrum, vibrates in synchrony with the pulsations of air pressure.

The simplest pattern of such pressure pulsations is generated for a "pure" tone, or sine wave. One important characteristic of the sine wave is its ***frequency.*** Frequency is the number of high to low variations in pressure, called ***cycles,*** that occur within a unit amount of time. The units we use to describe sound frequency are cycles per second, or ***Hertz (Hz).*** Waveforms of low and high frequency tones are illustrated in Figure 1.1.

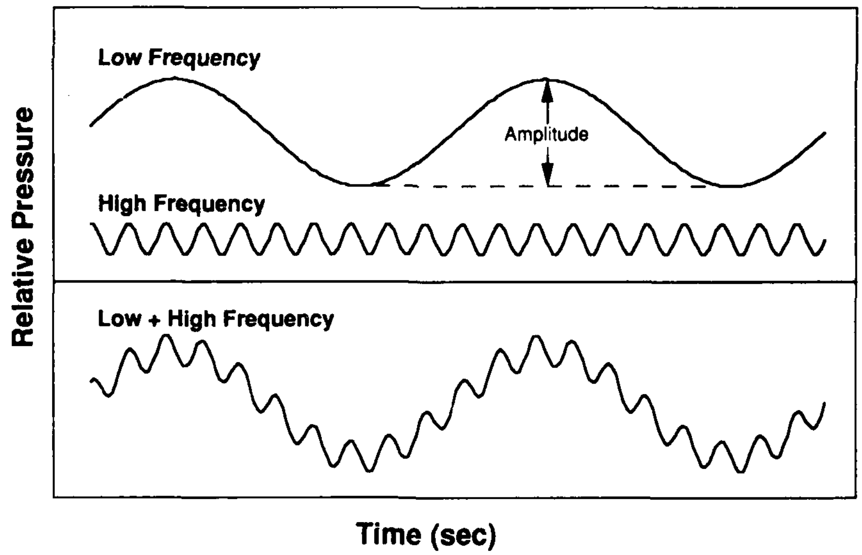


Figure 1.1. Changes in air pressure shown for two sound waves differing in frequency and amplitude (top). When added together (bottom), the two pure tones form a complex sound. (original figure)

Another important characteristic of pure tones is the degree of change from maximum to minimum pressure, which we call the ***amplitude*** or ***intensity***, also illustrated in Figure 1.1. Sound amplitude is usually measured in dynes per square centimeter, which is a measure of force per unit area. The human

auditory system is sensitive to an enormous range of variations in amplitude of a sound wave -- from about 1 to 10 billion. Thus, intensity is more conveniently specified by a logarithmic scale using units called ***decibels (dB)***. One dB = 20 log (p_1/p_0) where p_1 refers to the sound under consideration and p_0 is a standard reference (0.002 dynes per square centimeter). Table 1.1 shows some representative sounds on the dB scale.

Table 1.1
The Decibel Scale

dB	Example	Comment
0	Threshold of Hearing	
10	Normal Breathing	
20	Leaves Rustling	
30	Empty Office	
40	Residential Neighborhood at Night	
50	Quiet Restaurant	
60	Two-Person Conversation	
70	Busy Traffic	
80	Noisy Auto	
90	City Bus	
100	Subway Train	Prolonged Exposure Can Impair Hearing
120	Propeller Plane at Takeoff	
130	Machine-Gun Fire, Close Range	
140	Jet at Takeoff	Threshold of Pain
160	Wind Tunnel	

(Adapted from Sekuler & Blake, 1985)

Sine-wave tones are considered pure because we can describe any waveform as a combination of a set of sine waves each of which has a specific frequency and amplitude. This fact was initially demonstrated by Fourier. A sound comprised of more than a single sine wave is termed a ***complex sound***. Most of the sounds we hear are complex sounds. The bottom panel of Figure 1.1 shows how two sine waves of different frequencies can be combined to form a complex sound. Some typical complex sounds are shown in Figure 1.2. Here, we

see the same note from the musical scale played by three different musical instruments. Below each waveform is shown the amplitude of each frequency in the sound. That is, each complex sound was broken down into a set of sine waves of different frequencies using a method called Fourier analysis. The three instruments sound different because they contain different amplitude spectra (amplitudes as a function of frequency).

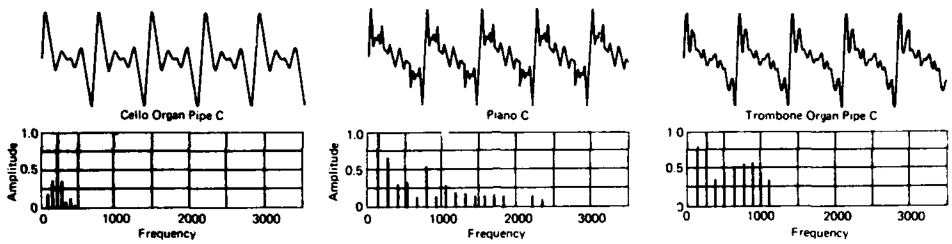


Figure 1.2. Amplitude spectra of a C note played on three different instruments. (from Fletcher, 1929)

Typically, the pitch we hear in a complex sound corresponds to the pitch of the lowest frequency component of that sound. This component is called the ***fundamental frequency***. Frequency components higher than the fundamental are called ***harmonics***, and these harmonics affect the quality or the ***timbre*** of the sound. Two musical instruments, say a trumpet and piano, playing the same note will generate the same fundamental. However, their higher frequency components, or harmonics will differ, as illustrated in Figure 1.2. These harmonics produce the characteristic differences in quality between different instruments. If we were to remove the harmonics, leaving only the fundamental, a trumpet and a piano playing the same note would sound identical.

Frequency and Intensity Relations to Perception

How do the physical properties of sound relate to our perceptions? First, consider the range of frequencies over which we are sensitive. The lower curve in Figure 1.3 shows how absolute threshold varies with sound frequency for a young adult.

The range over which sounds can be detected is from about 20 to 20,000 Hz. As you can see, we are most sensitive to sounds between 500 and 5,000 Hz. These are also the frequencies of human speech. The frequency range

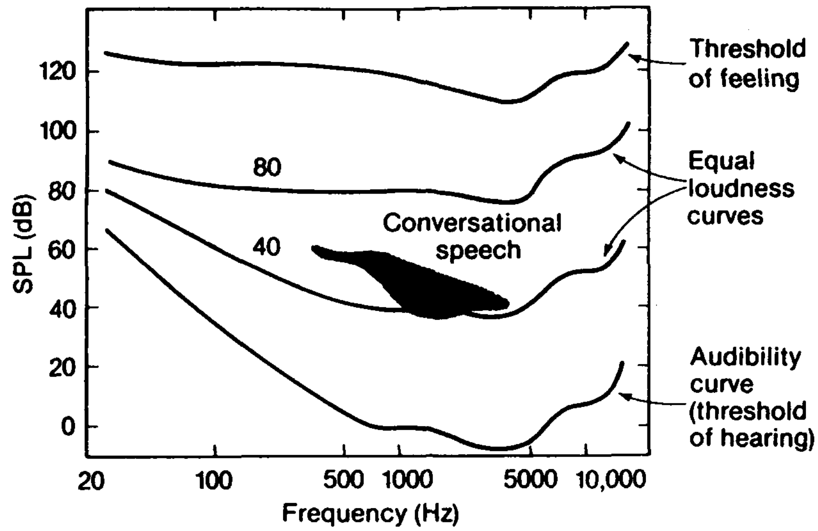


Figure 1.3. Variation of absolute threshold with sound frequency for a young adult. (from Fletcher & Munson, 1933)

recommended for aircraft warning signals is 250 to 4,000 Hz (FAA voluntary guidelines based on FAA RD-81/38,II, page 91). Our sensitivity declines sharply, i.e., the threshold increases, for higher and lower frequencies. What this means is that sounds of different frequencies require different amounts of energy to be loud enough to be heard.

It is interesting to consider our absolute sensitivity under optimal conditions. At about 2,500 Hz, we are so sensitive that we can detect a sound that moves the eardrum less than the diameter of a hydrogen molecule (Békésy & Rosenblith, 1951). In fact, if we were any more sensitive, we would hear air molecules hitting our eardrums and blood moving through our head.

To make the frequency scale a little more intuitive, consider that the range on a piano is from about 27.5 Hz to about 4,186 Hz. Middle C is 262 Hz. As sound frequency is increased from 20 to 20,000 Hz, we perceive an increase in ***pitch***. It is important to note though that our perception of pitch does not increase in exact correspondence to increases in frequency.

As we increase the amplitude, or physical intensity, of a particular frequency, its loudness increases. Loudness is a perceptual attribute referring to our subjective experience of the intensity of a sound, however, not a physical property of the sound. To measure the relative loudness of a sound, researchers typically present a tone of a particular frequency at a fixed intensity and then ask subjects to increase or decrease the intensity of another tone until it matches the loudness of the standard. This is repeated for many different frequencies to yield an ***equiloudness contour***. Figure 1.3 shows equiloudness contours for

standards of 40 and 80 dB above threshold. Note that the shape of the contour changes with increasing intensity. That is, the increase in the loudness of a sound with increasing intensity occurs at different rates for different frequencies. Thus, we are much more sensitive to intermediate frequencies of sound than to extremes in frequency. However, with loud sounds, indicated by higher intensity standards in Figure 1.3, this difference in our sensitivity to various frequencies decreases.

Sensitivity to loudness depends on the sound frequency in a way that changes with the level of sound intensity. You have probably experienced this phenomenon when listening to music. Listen to the same piece of music at high and low volumes. Attend to how the bass and treble become much more noticeable at the higher volume. Some high-fidelity systems compensate for this change by providing a loudness control that can boost the bass and treble at low volume. The fact that the loudness of a tone depends not only on its intensity but also on its frequency is a further illustration that physical and perceptual descriptions are not identical.

While pitch depends on frequency, as mentioned, it also depends on intensity. When we increase the intensity of a low frequency sound, its pitch decreases. When we increase the intensity of a high frequency sound, its pitch increases.

The Effects of Aging

The frequency range for an individual observer is commonly measured by audiologists and is known as an ***audiogram***. Figure 1.3 showed that the frequency sensitivity of a young adult ranged from about 20 to 20,000 Hz. This range diminishes with increasing age, however, so that few people over age 30 can hear above approximately 15,000 Hz. By age 50 the high frequency limit is about 12,000 Hz and by age 70 it is about 6,000 Hz (Davis & Silverman, 1960). This loss with increasing age is known as ***presbycusis***, and is usually greater in men than in women.

The cause of presbycusis is not known. As with all phenomena of aging, there are large individual differences in the magnitude of high frequency hearing loss. One possibility is that changes in vasculature with increasing age limit the blood supply to sensitive neural processes in the ear. Another possibility is that there is some cumulative pathology that occurs with age. For example, cigarette smokers have a greater age-related loss in sensitivity than nonsmokers (Zelman, 1973) and this may be due to the interfering effects of nicotine on blood circulation. There are other possibilities, but perhaps the most important to consider is the cumulative effect of sound exposure.

Effects of Exposure

Sudden loud noises have been known to cause hearing losses. This is a common problem for military personnel exposed to gun shots. Even a small firecracker can cause a permanent loss in hearing under some conditions (Ward & Glorig,1961).

Exposure to continuous sound is common in modern industrial societies. Even when the sounds are not sufficiently intense to cause immediate damage, continuous exposure may produce loss of hearing, especially for high frequencies. Unprotected workers on assembly lines or airports have hearing losses that are correlated with the amount of time on the job (Taylor, 1965). Similar studies have shown deleterious effects of attending loud rock concerts.

The potentially damaging effects of sound exposure on hearing depend on both the intensity and duration of the sounds. Thus, cumulative exposure to sound over the life span might be related to presbycusis.

Sound Localization

The separated locations of our ears allows us to judge the source of a sound. We use incoming sound from a single source to localize sounds in space in two different ways. To begin with, suppose a tone above 1,200 Hz is sounded directly to your right, as illustrated in Figure 1.4.

The intensity of high frequency sounds will be less in the left ear than the right because your head blocks the sounds before they reach your left ear. This ***intensity difference*** only exists for sounds above 1,200 Hz, however. At lower frequencies, sound can travel around your head without any significant reduction in intensity.

Whenever a sound travels farther to reach one ear or the other, a ***time difference*** exists between the arrival of the sound at each ear. Thus, if the sound source is closer to one ear, the pulsations in air pressure will hit that ear first and the other a bit later. We can use a time difference as small as 10 microseconds between our two ears to localize a sound source (Durlach & Colburn, 1978), but this information is only useful for low frequency sounds. Thus, localization of high frequency sounds depends primarily on interaural intensity differences, but low frequency sounds are localized by interaural time differences.

Habituation and Adaptation

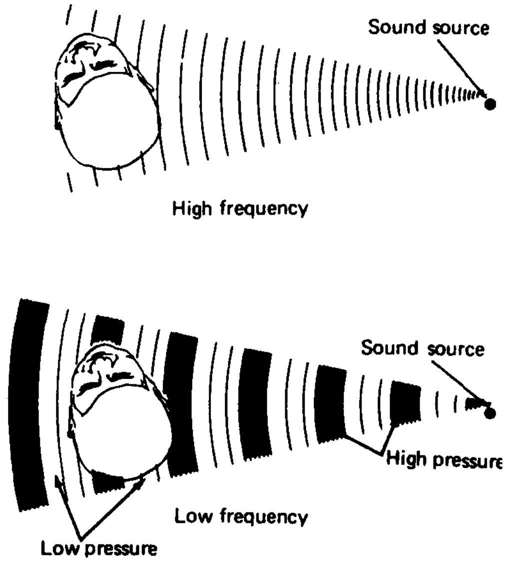


Figure 1.4. High and low frequency sound waves emanating from a source to the right of a person's head. (from Werner & Schlesinger, 1991)

Our ability to detect sounds is not static, but rather changes as a sound is repeatedly presented. This can be due to ***adaptation***, a physiological change in sensitivity of the auditory system following exposure to sounds. However, changes in the ability to detect sounds need not occur for us to "tune out" sounds around us. When a stimulus is repeatedly presented, there is a tendency to decrease responsiveness over time. For example, when sitting in a room we may notice a fan when it is first turned on, but over time the noise of the fan is not noticeable at all. This is called ***habituation***, a decrease in response or of noticing the sound that cannot be attributed to fatigue or adaptation. To distinguish between adaptation and habituation, the same tone might suddenly be reduced in intensity. If the response is due to habituation, there may be a recovery of response even though the stimulus is weaker.

The importance of habituation is clear when an individual must engage in a task that involves attending to repetitive stimuli. There is a natural tendency to tune out what is repeated and renew attention to what is novel. Tuning out what is repeated, and presumably irrelevant, keeps the sensory channels open to process new information. Habituation and adaptation phenomena are not limited to detecting auditory stimuli, but they can be demonstrated for any of the senses.

Ambient Noise (Masking)

Detection of pure tones is affected by background noise. We require more intense tones for detection in the presence of background noise, and the shape of the frequency sensitivity curve changes with the characteristics of the ambient noise. The experience of detecting sounds in the presence of background noise is a familiar one. In the laboratory we call the sound that an individual is trying to detect the target, and the sound that is interfering with detection the ***masking*** stimulus. Not surprisingly, the effectiveness of a masking

stimulus increases with its intensity. This corresponds to our experience in which we must speak more loudly to be heard as the sounds around us increase in loudness. Perhaps not so intuitive, however, are the results of masking studies which show that masking sounds do not affect all tones equally, but rather act selectively to reduce sensitivity for tones of the same and somewhat higher frequencies than the mask (Zwicker, 1958).

There are some conditions in which having two ears makes it possible to reduce the effects of masking stimuli. To demonstrate this effect, sounds are played separately to the two ears by use of headphones. Suppose that a tone is delivered to the right ear and it becomes inaudible when masking noise is delivered to that same ear. Now, if the same noise stimulus (without the tone) is played to the other ear, the tone will become audible again. It is as though the stimuli to both ears can be separated from the target that is presented to only one ear. This is known as ***binaural unmasking***.

Binaural unmasking is probably one factor that helps an individual to focus on one set of sounds in the presence of others. This is a familiar experience at parties, in which you can listen to one conversation while tuning out conversations in the background. If your name happens to be mentioned in another conversation, however, you may find yourself unable to resist switching the conversation to which you are listening. This is known as the ***cocktail party phenomenon*** and it underscores our ability to monitor incoming information that we are not actively processing.

Chapter 2

Basic Visual Processes

by John S. Werner, Ph.D., University of Colorado at Boulder

Vision is our dominant sensory channel, not only in guiding aircraft, but also in most tasks of everyday life. For example, we can recognize people in several ways -- by their appearance, their voice, or perhaps even their odor. When we rearrange stimuli in the laboratory so that what one hears or feels conflicts with what one sees, subjects consistently choose responses based on what they saw rather than on what their other senses told them (Welch and Warren, 1980). Most of us apparently accept the idea that "seeing is believing."

Physical Properties of Light

Light is a form of electromagnetic energy that is emitted from a source in small, indivisible packets called ***quanta*** (or ***photons***). A quantum is the smallest unit of light. As with sound energy, the movement of light energy through space is in a

sinusoidal pattern. Sound waves were described in terms of their frequency, but light waves are more commonly described in terms of the length of the waves (i.e., the distance between two successive peaks). This description is equivalent to one based on frequency because wavelength and frequency are inversely related. Figure 2.1 illustrates two waves differing in their length. As can be seen

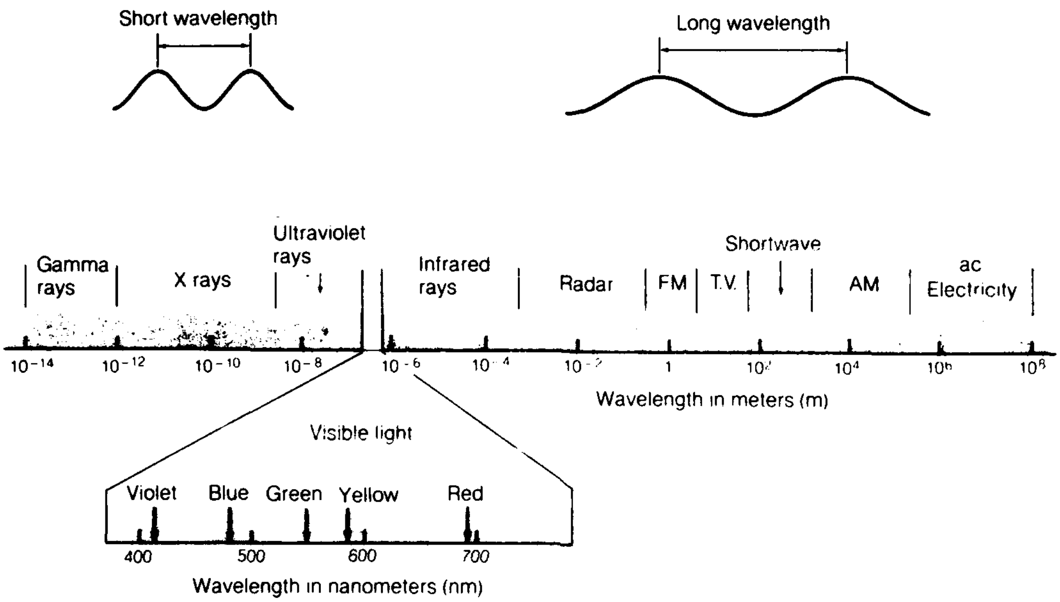


Figure 2.1 Regions of the electromagnetic spectrum and their corresponding wavelengths. (from Coren, Porec & Ward, 1984)

in the figure, the electromagnetic spectrum encompasses a wide range, but our eyes are sensitive only to a small band of radiation which we perceive as light. Normally, we can see quanta with wavelengths between about 400 and 700 ***nanometers*** (nm; 1 nm is one billionth of a meter). Thus, the two major physical variables for discussing light are quanta and wavelength. The number of quanta falling on an object describes the light intensity, whereas the wavelength tells us where the quanta lie in the spectrum. Most naturally occurring light sources emit quanta of many wavelengths (or a broadband of the spectrum), but in a laboratory, we use specialized instruments that emit only a narrow band of the spectrum called ***monochromatic*** lights. If a person with normal color vision were to view monochromatic lights in a dark room, the appearance would be violet at 400 nm, blue at 470 nm, green at 550 nm, yellow at 570 nm, and red at about 680 nm. Note that this description is for one set of conditions; later we will illustrate how the appearance can change for the same monochromatic lights when viewed under other conditions.

Figure 2.2 shows the distribution of energy for some familiar light sources, fluorescent lamps. The four different curves show four different types of lamp.

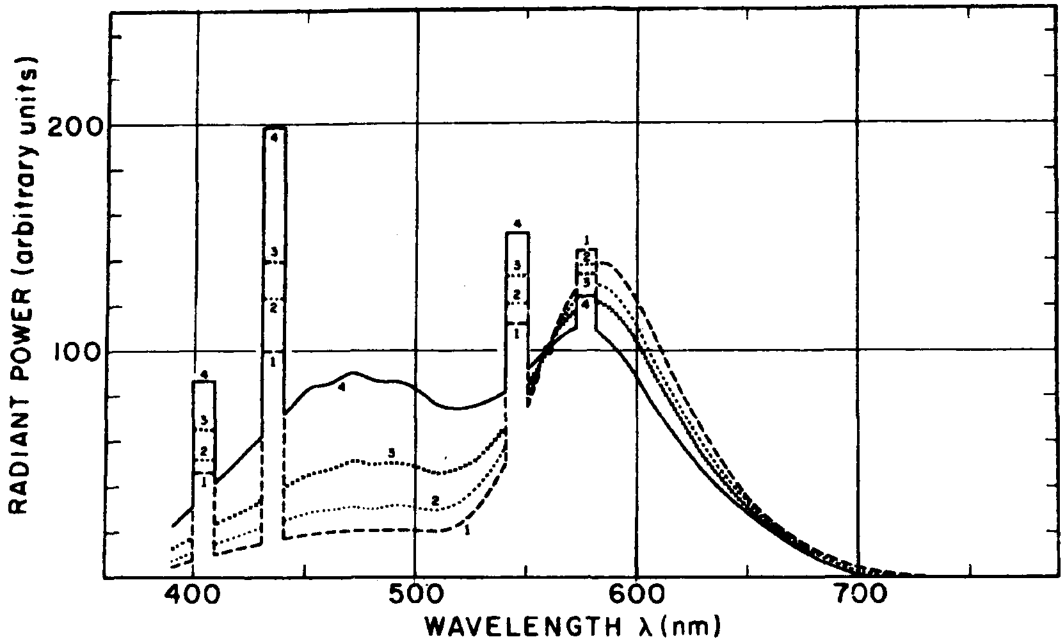


Figure 2.2 **Relative energy of fluorescent lamps plotted as a function of wavelength: 1 = standard warm white, 2 = white, 3 = standard cool white, 4 = daylight. (from Wyszecki & Stiles, 1982)**

While they all may be called "white," they differ in their relative distribution c. energy. They also appear different in their color although this is not always noticed unless they are placed side-by-side. Variations in the intensity and spectral distribution of energy can sometimes be quite large without affecting our color perception. Indeed, Figure 2.3 shows the energy of sunlight plotted as a function of wavelength for a surface facing away from the sun or toward the sun. If these two light distributions were placed side-by-side you would say that one is bluish and the other yellowish, but if either one was used to illuminate a whole scene by itself, you would most likely call this illuminant white and objects would appear to have their usual color. Objects usually do not change their color with these changes in the source of illumination. This perceptual phenomenon is called ***color constancy***.

When light travels from one medium to another, several things can happen. First, some or all of the quanta can be lost by ***absorption*** and the energy in the absorbed quanta is converted into heat or chemical energy. Second, when striking another medium some or all of the quanta can bounce back into the initial medium, a familiar phenomenon known as ***reflection***. Third, the light can

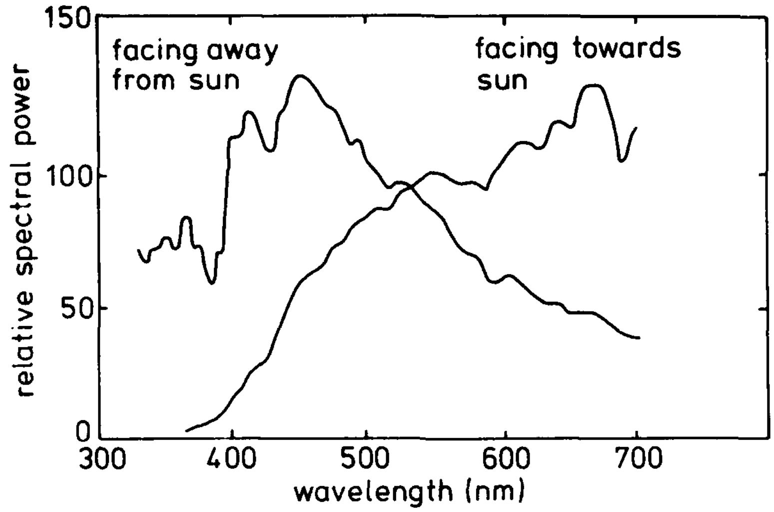


Figure 2.3. **Sunlight energy plotted as a function of wavelength for a surface facing away from (30° solar altitude) or toward the sun (8° solar altitude). (from Walraven et al., 1990)**

be transmitted, or move forward, from one medium to another, but in doing so the path may change somewhat; that is, the rays of light will be bent by ***refraction***. The extent to which each of these phenomena will occur depends upon the medium that the light is striking, and the angle of incidence between the light rays and the medium.

Absorption, reflection, and refraction all occur at the various structures in the eye. It is, therefore, important to consider these phenomena in attempting to understand the formation of optical images in the eye.

The Eye

Figure 2.4 is a diagram of the human eye. The eyeball is surrounded by a tough, white tissue called the ***sclera***, which becomes the clear ***cornea*** at the front. Light that passes through the cornea continues on through the ***pupil***, a hole formed by a ring of muscles called the ***iris***. It is the outer, pigmented layer of the iris that gives our eyes their color.

Contraction and expansion of the iris opens or closes the pupil to adjust the amount of light entering the eye. Light then passes through the ***lens*** and strikes the ***retina***, several layers of cells at the back of the eye. The retina includes ***receptors*** that convert energy in absorbed quanta into neural signals. One part of the retina, called the ***fovea***, contains the highest number of receptors per unit area. When we want to look at an object, or fixate it, we move our head and

eyes so that the light will travel along the visual axis and the image of the object will fall on the fovea.

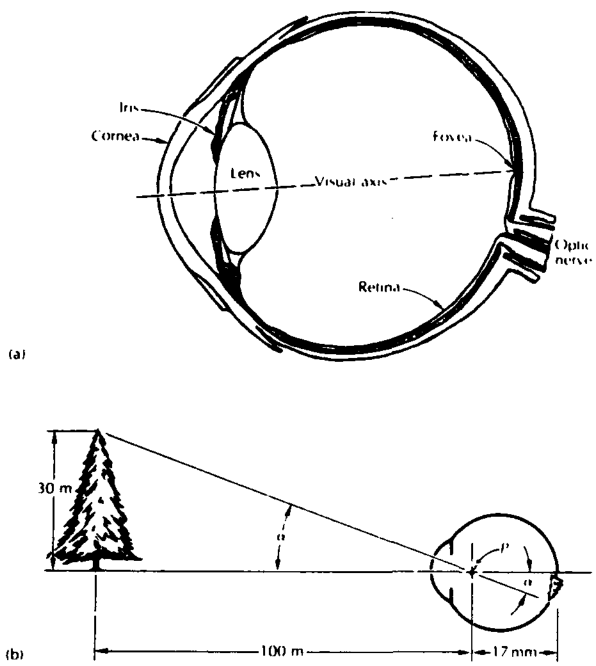


Figure 2.4. Cross section of the human eye; visual angle. (from Cornsweet, 1970)

The sizes of visual stimuli are often specified in terms of the region of the retina that they subtend (cover). This concept is illustrated in panel (b) of Figure 2.4. Consider what happens when we look at an object, say a tree (Figure 2.4). Imagine the tree as many points of light, and we are looking at the light coming from the top of the tree. When we focus on the tree, our cornea and lens bend the light so that an image of the tree is formed at the back of the eye, much as an image is made on photographic film by a camera. Note that the optics of the eye bend the light so that the image of the tree on the retina is upside down and reversed left to right. The area of the retina covered by the image is called the visual angle, which is measured in degrees. The angle depends on the object's size and distance from us. In Figure 2.4, we can deduce that smaller and smaller trees at closer and closer distances could all subtend the same visual angle. The same principles hold for two equally sized objects at differing distances; they will produce different visual angles and appear as different sizes. This relation is such that as the distance of an object from the eye doubles, the size of the image produced by the object is halved. Artists use this information to create an illusion of three-dimensional space on a flat surface by making background figures smaller than foreground figures.

The visual angle x is calculated by: arctan (size/distance), and is specified in degrees. (Note that the distance is between the object and the cornea, plus the distance between the cornea and point 'p' in Figure 2.4. The latter value is seven mm.) By definition, one degree equals 60 minutes of arc, and one minute of arc equals 60 seconds of arc. A rough rule of thumb (no pun intended) is that the visual angle 'x' of your thumb nail at arms length is about 2°.

An eye that properly focuses distant objects on the retina is said to be ***emmetropic***. Howeve. as Figure 2.5 illustrates, some individuals have an eyeball that is abnormally short or the optics of their eye do not sufficiently refract the incoming light with the result that the light is focused behind the retina. This condition is called ***hypermetropia*** or farsightedness. Other individuals may have an eye that is too long or optics that refract the light from distant objects too much with the result that the object is imaged in front of the retina. This condition is known as ***myopia*** or nearsightedness. In both hypermetropia and myopia, the image falling on the retina is not properly focused and vision may be blurred. Fortunately, this problem can be corrected by prescribing spectacles or contact lenses that cause distant images to be focused on the retina.

Accommodation

At any one time, the eye can focus objects clearly only if those objects fall within a limited range of distance. To look at close objects we require more bending of the light to properly focus the image on the retina. In humans, this is accomplished by a somewhat flexible lens in the eye. The lens is attached to muscles that can be contracted or relaxed to change the lens curvature. When the shape is changed, the light will be refracted or bent differently, a process known as ***accommodation***.

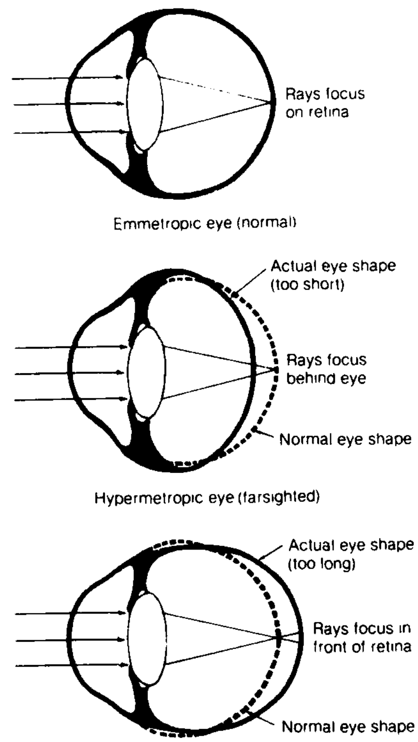


Figure 2.5. **Image formation in emmetropic (normal), hypermetropic, and myopic eyes. (from Coren, Porac, & Ward, 1984)**

It is not clear what triggers the eye to change its state of accommodation, but one likely source is a defocused image. Since shifts in fixation from far to near objects will be associated with some image blur, accommodation will occur. The reaction time for accommodation is about 360 milliseconds (Campbell & Westheimer, 1960). Although this is a short reaction time, it is nevertheless long enough to produce noticeable blur when shifting focus from a display panel or head-up display (HUD) to a distant object, or vice versa. It may be noted that the need to accommodate to HUD symbology is theoretically unexpected because it is produced by optically

collimated virtual images; however, collimated images do not necessarily lead to focus at optical infinity (Iavecchia, Iavecchia & Roscoe, 1988).

Aging and Presbyopia

The flexibility of the eye lens decreases with age and thereby limits the ability to accommodate, both in terms of the amount of change in the lens and the time required to respond to changes that occur when shifting fixation from far to near objects (Weale, 1982). The loss in accommodative ability, known as ***presbyopia***, is often quantified in terms of the near point, or the closest distance at which an object can be seen without blur. As illustrated by Figure 2.6, the near point increases with advancing age. By about age 40, the near point is such that reading can only be accomplished when the print is held at some distance or if reading glasses are used.

Some individuals require one lens correction for their distance vision and a different correction for their presbyopia. This can be accomplished by bifocal lenses -- lenses which require the individual to look through different parts in order to properly focus near and far objects.

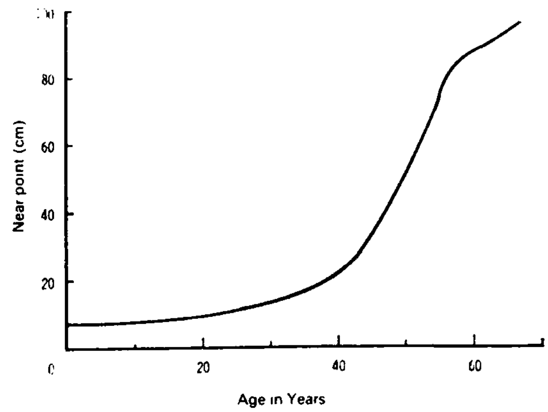


Figure 2.6. Near point plotted as a function of age. (from Helps, 1973)

Ocular Media Transmission and Aging

The various optical components of the eye -- the ocular media -- shown in Figures 2.4 and 2.5 are not completely transparent. The lens of the eye, in particular, has a yellowish color. It absorbs quite strongly at the short wavelengths of the visible spectrum (around 400 to 450 nm) and even more strongly in the ultraviolet portion of the spectrum from 300 to 400 nm. This is illustrated by Figure 2.7 which shows optical density plotted as a function of wavelength. Optical density refers to the log of the reciprocal of transmission and can be thought of as the log of the absorption. Thus, optical density 2.0 refers to ten times greater absorption than optical density 1.0.

Figure 2.8 shows the variation in ocular media density (at 400 nm) as a function of advancing age. One can see that at each age there is a great deal of individual variation, about 1 log unit or a factor of 10-to-1. In addition, the

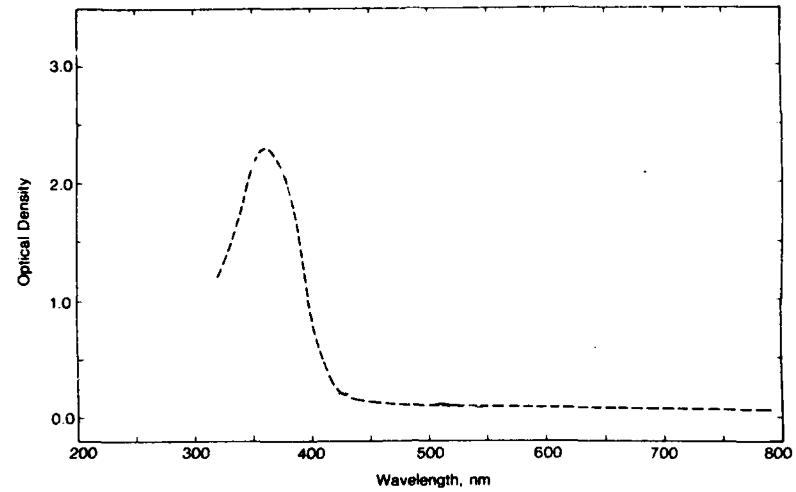


Figure 2.7. Optical density of the human lens plotted as a function of wavelength. (data from Boettner, 1967, original figure)

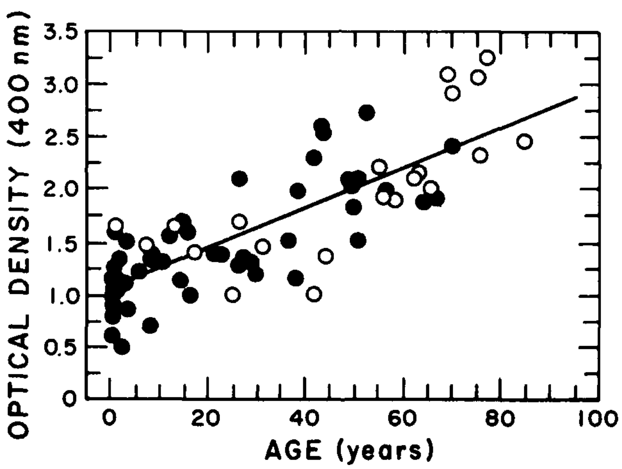


Figure 2.8. Optical density of human ocular media at 400 nm plotted as a function of age. (from Werner, Peterzell & Scheetz, 1990)

optical density of the lens increases markedly with advancing age. It can be deduced from the solid line fit to the data that the **average 70-year-old eye transmits about 22 times less light at 400 nm (1.34 optical density difference) than does the eye of the average 1-month-old infant.** This difference between young and old diminishes with increasing wavelength.

Because the lens increases its absorption with age, the visual stimulus arriving at the receptors will be less intense with age. In addition, for stimuli with a broad spectrum of wavelengths, there will be a change in the relative distribution of light energy because the short wavelengths will be attenuated

more than middle or long wavelengths. Since the stimulus at the retina is changing with age, **there will be age-related decreases in the ability to detect short wavelengths of light.** The amount of light absorbed by the lens will also directly influence our ability to discriminate short wavelengths (blue hues). Thus, the large range of individual variation in the lens, leads to large individual differences in discrimination of blue hues and in how a specific blue light will appear to different observers.

While an increase in the absorption of light with advancing age is considered normal, some individuals experience an excessive change which leads to a lens opacity known as a ***cataract***. A cataractous lens severely impairs vision and is typically treated by surgical removal and implantation of a plastic, artificial lens. These artificial lenses eliminate the ability to accommodate, but in most cases of cataract the individual is above about 55 or 60 years of age and has lost this ability anyway.

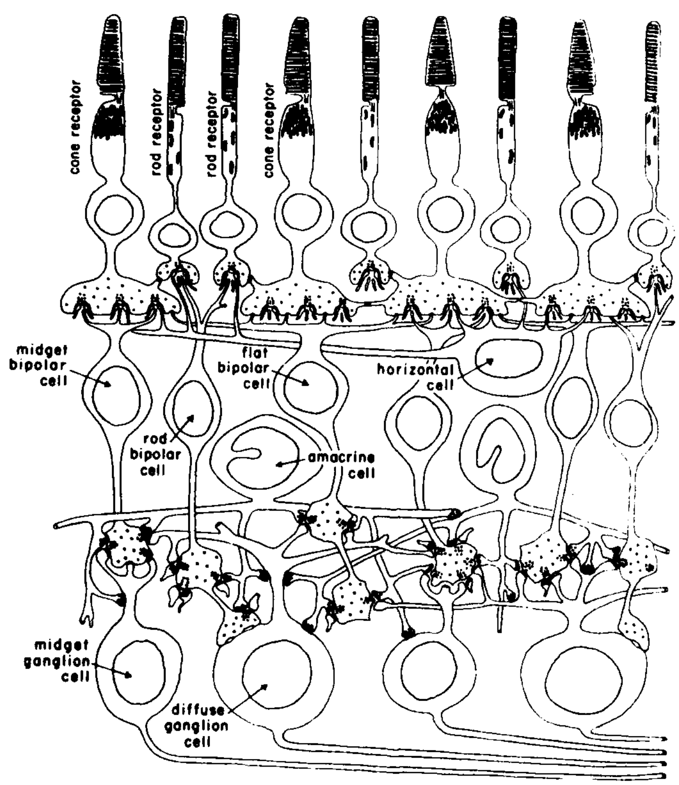


Figure 2.9. Various cell types in the primate retina. (original figure from Dowling & Boycott, 1966; modified by Wyszecki & Stiles, 1982)

Rods and Cones

Figure 2.9 shows the various cell layers of the retina. Quanta falling on the retina are absorbed by photopigments contained in the visual receptors. Energy contained in an absorbed quantum changes the structure of the photopigment which causes the receptor to respond. These responses are passed along to other cells in the retina. In this diagram, light enters from the bottom of the picture and before it reaches the receptors it must travel through the different cell layers. This does not affect the image, however, since these other cells are essentially transparent. The human retina contains two types of visual receptors, the rods and ***cones***, so named because of their different shapes.

Variation with Retinal Eccentricity

There are approximately 6 million cones and about 125 million rods in the human retina. These receptor cells are not evenly distributed across the retina, as shown in Figure 2.10. The cones are most densely packed in the fovea. To look at something directly or fixate on it, we turn our eyes so that the object's image falls directly on the fovea. This is advantageous because the fovea contains the greatest number of cones, providing us with our best ***visual acuity***, or ability to see fine details. Outside the fovea, where the density of the cones decreases, there is a corresponding decrease in visual acuity. The density of rods is greatest about 20° from the fovea and decreases toward the periphery. The periphery has many more rods than cones, but a careful reading of the figure shows that there are as many as 7,500 cones per square mm even in the peripheral retina.

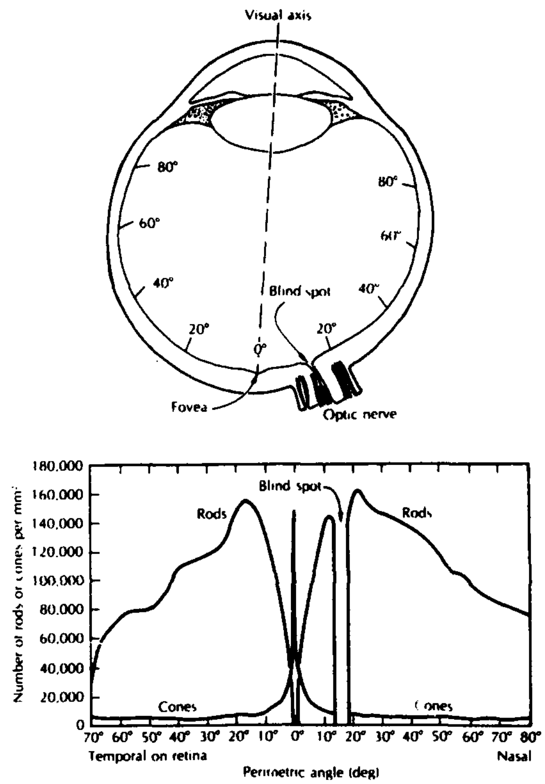


Figure 2.10. **The number of rods and cones plotted as a function of retinal eccentricity. (data from Osterberg, 1935; figure from Cornsweet, 1970)**

When light falls on the rods and cones, they send signals to other retinal cells, the horizontal, bipolar, and amacrine cells located in different retinal layers (Figure 2.9). These cells organize incoming information from receptors in complex ways. For example, one of these cells can receive information from many receptors as well as from other retinal cells. Then these cells send their information on to ganglion cells, which can further modify and reorganize the neural information. The activity of these ganglion cells is sent to the brain along neural fibers called axons. Thus, the only information that our brain can process must be coded in the signals from the ganglion cells. The interactions among the different retinal cell types provide the physiological basis for many important perceptual phenomena.

The axons of ganglion cells form a bundle of approximately one million fibers called the optic nerve. These fibers leave the eyeball in the region termed the ***optic disc***. Because this area is devoid of receptors, it is called the ***blind spot***. As can be seen in Figure 2.10, the blind spot is located at about 15° on the side of the nose (or nasal retina) from the fovea.

As a practical matter, one can now see why FAA guidelines (see RD-81/38,II, page 40) stress the importance of **placing master visual alerts within 15° of each pilot's normal line of sight** as illustrated by Figure 2.11. This is the area of the visual field with best visual acuity and typically the center of attention. By

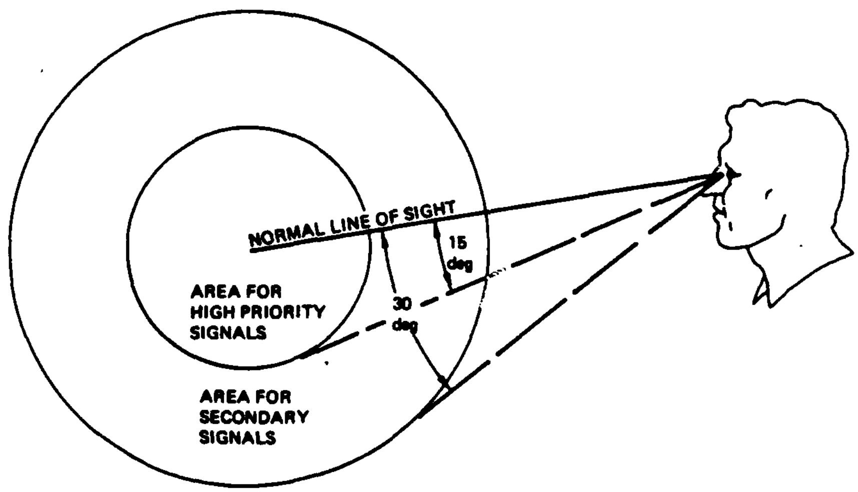


Figure 2.11. **Recommended placement of visual alert and other high priority signals relative to the line of sight. (from DOT/FAA/RD-81/38,II)**

placing high priority signals in this area, they will be detected more quickly than if they are placed more peripherally.

Spectral Sensitivity

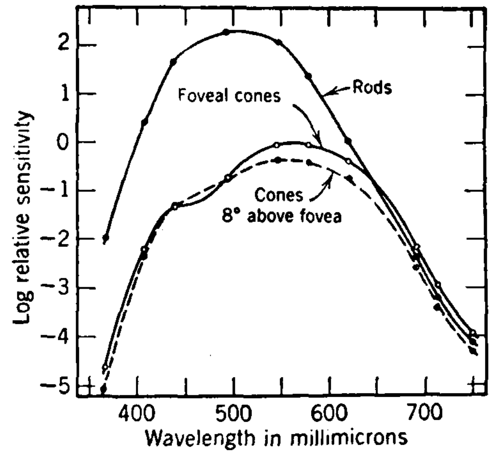


Figure 2.12. Log relative sensitivity plotted separately for rods and cones as a function of wavelength. (Data from Wald, 1945; figure from Judd, 1951)

The functional difference between rods and cones was discovered in 1825, when the Czech medical doctor Purkinje realized that he was most sensitive to a part of the spectrum in complete darkness that was different from the part he was most sensitive to in daylight. From this, he hypothesized the existence of different receptors for day (***photopic***) vision and night (***scotopic***) vision. Shortly after this, a German biologist named Schultze described two types of receptors in the retina which he named rods and cones based on their shapes. He noted that rods were the main type of receptor in animals active at night and cones predominated in animals active during the day. From this, he concluded that rods are the receptors of scotopic or "night" vision and cones the receptors for photopic or "daylight" vision.

Rods and cones differ in their sensitivity to different wavelengths of light, or their ***spectral sensitivity***. If we measure spectral sensitivity with light focused on the retina where the rods are most numerous, the maximal sensitivity will be at about 505 nm. The top curve in Figure 2.12 shows ***scotopic spectral sensitivity***, or the sensitivity of rods to different wavelengths. The shape of this curve is due to the fact that the photopigment contained in rods absorbs some wavelengths better than others.

Under scotopic conditions, we do not perceive hue -- the chromatic quality in colors that we identify with names such as red, green, blue, and yellow. If we observed lights of different wavelengths emitting the same numbers of quanta, light at 510 nm would appear brighter to us than other wavelengths because of our greater sensitivity to it, but all the wavelengths would appear to have the same color under scotopic or dark-adapted conditions.

If we measure spectral sensitivity for cones by focusing light directly onto the fovea where there are virtually no rods, we see that cone sensitivity is dramatically lower than for the rods -- as much as a thousand times lower at

some wavelengths. The wavelength of maximal sensitivity for the cones (555 nm) is different than for the rods. This ***photopic spectral sensitivity*** is shown in Figure 2.12. Not only do the cones differ from the rods in their spectral sensitivity, but they produce different perceptual experiences. Under photopic or daylight conditions, we can see different hues as wavelength varies. Thus, perception of hue is dependent on cone receptors.

Luminance

Let us briefly digress to consider a practical implication of the spectral sensitivity functions. We have seen that lights can be specified in terms of the number of quanta emitted at various wavelengths of the visible spectrum. However, because the eye is not equally sensitive to all wavelengths, the specification of the intensity of a light in terms of a purely physical metric does very little to describe its effectiveness as a stimulus for vision. For this reason, the International Commission on Illumination (Commission Internationale de l'eclairage, CIE) has developed a system of specifying the intensity of the stimulus weighted according to the spectral sensitivity of the human observer. The spectral sensitivity function used by the CIE is called the standard observer's visibility function or V_λ when specifying lights under photopic conditions and V'_λ when specifying lights viewed under scotopic conditions. ***Luminance***, the intensity of light per unit area reflected from a surface toward the eye, is thus defined as:

$$K\int E_\lambda V_\lambda d\lambda$$

where E_λ is the radiant energy contained in wavelength interval $d\lambda$ and V_λ is the relative photopic spectral sensitivity function for the standard observer. For scotopic conditions, the same formula applies except that V'_λ is used instead of V_λ. The K is related to the units in which luminance is specified, the most common in current usage being the candela per square meter (cd/m^2). In the literature, one may find luminance specified in different units by different investigators. Conversion factors needed to compare the various studies are tabled by Wyszecki and Stiles (1982).

There are a few points to note about luminance specifications. First, there is no subjectivity inherent in the measurement of luminance. One simply measures the energy at each wavelength and multiplies this value by the relative sensitivity of the standard observer at that wavelength. Alternatively, one may directly measure luminance with a meter -- a meter that has been calibrated to have the sensitivity of the CIE standard observer. Second, while there is no subjectivity in the measurement of luminance, it was the original intent of the CIE to develop a metric that would be closely related to the brightness or subjective intensity

of a visual stimulus. As we shall see, the brightness of a stimulus depends on many variables such as the preceding or surrounding illumination. These variables are not taken into account in specifying luminance. Thus, the luminance of a stimulus is often of no value in specifying brightness (Kinney, 1983). The term luminance should be reserved for the specification of light intensity, and the term brightness should be reserved for a description of the appearance of a stimulus.

Dark Adaptation

Most of us have groped around in a dark movie theater until our eyes adjusted to the dim level of illumination. This process is called ***dark adaptation***, and it occurs, in part, because our receptors need time to achieve their maximum sensitivity, or minimum threshold. If we were to measure the minimum amount of light required to see at various times, i.e., our threshold after we entered a darkened room, we could plot a dark adaptation curve such as that shown in Figure 2.13 (reprinted by permission from C.H. Graham, Ed. ***Vision and Visual Perception***, © John Wiley & Sons, Inc. New York, NY, 1965, p. 75). This curve indicates that the eye becomes progressively more sensitive in the dark, but notice that the curve has two distinct phases. The first phase, which lasts about seven minutes, is attributed to the cone system, and the second phase, to the rod system. When we first enter the dark our cones are more sensitive than the rods, but after about seven minutes, the rods become more sensitive.

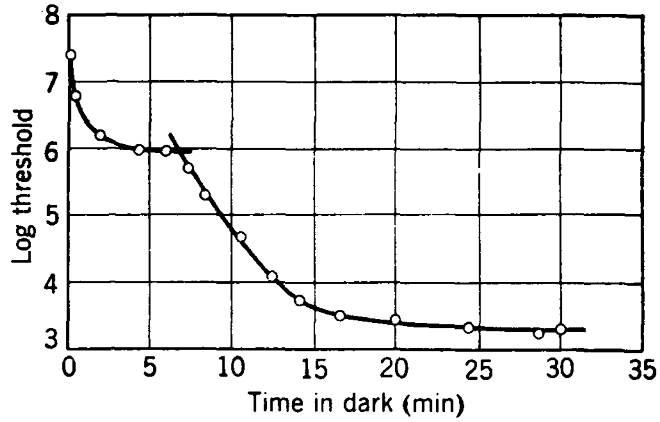


Figure 2.13. Threshold decrease during adaptation to darkness showing that cones (top branch) and rods (bottom branch) adapt at different rates. (from Graham, 1965a)

What explains the greater sensitivity of rods over cones in a dark theater? Part of the answer is related to the fact that there are many more rods than cones. Second, because the rods contain more photopigment than cones, they absorb more quanta. To consider the third explanation for the difference in scotopic and photopic sensitivity, we must look at the connections of rods and cones to other neural elements in the retina. Several cones are often connected to a single bipolar cell. This is termed **convergence** because the signals from several

cones come together at one cell. The more receptors converging on a single cell, the greater chances are of activating that cell.

Sensitivity/Resolution Trade-Off

A dim light that produces a weak signal in many rods has a greater chance of being detected because many rods summate their signals on another cell. Their combined effects can produce a signal strong enough for visual detection. Detection of a weak signal by cones is less likely because their spatial summation of signals occurs over much smaller regions than rods. Convergence, a structural property of many neural-sensory systems, thus enhances sensitivity. (Signals in receptors can also be added together over time, a process known as temporal summation, and this occurs over longer durations for rods than for cones.)

While it may seem advantageous to summate visual signals over a wide region of the retina to enhance sensitivity, it should be noted that this is associated with a loss of resolution or acuity. That is, whenever signals are combined, information about which receptors generated the signals is lost. Conversely, if information from receptors is separated, there is a greater possibility of localizing which receptors are activated and thereby resolving the locus of stimulation. Thus, there is a trade-off between sensitivity and resolution. Because rods pool their signals over larger retinal regions than cones, they enhance sensitivity at the cost of spatial resolution. Cones, on the other hand, summate information over small regions of retina and favor high resolution at the expense of sensitivity.

Visual acuity, or resolution, is often defined in terms of the smallest detail that an observer can see. This is measured by the familiar eye chart with varying letter sizes viewed at a fixed distance. Visual acuity tested with such a chart is defined by the smallest letter that can be read. When an individual has, for example, an acuity of 20/40 or 0.5 it means that at a distance of 20 feet, the individual just resolves a gap in a letter that would subtend 1 minute of arc at a distance of 40 feet (see Riggs, 1965 for other details). In many states, a person is legally blind if visual acuity is 20/400 or worse.

Visual acuity varies with luminance, as shown in Figure 2.14. In the scotopic range, visual acuity is dependent on rods and is very poor. As light intensity increases into the photopic range, visual acuity is more dependent on cones and dramatically improves. Note, however, that even after cones "take over," visual acuity continues to vary with light intensity. The data in Figure 2.14 represent more or less ideal conditions. When a stimulus is moving or the display is vibrating (as in turbulence), visual acuity may be considerably reduced.

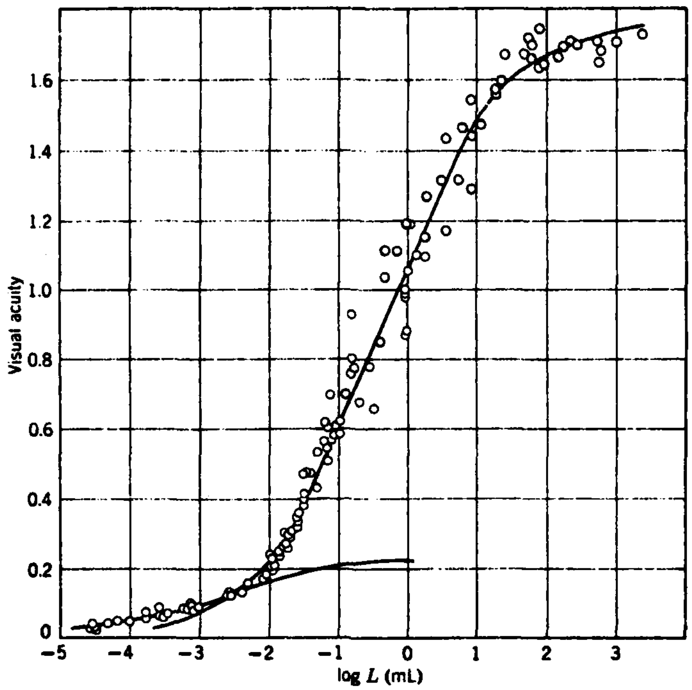


Figure 2.14. Visual acuity plotted as a function of log luminance. (from Hecht, 1934)

Damage Thresholds

The human visual system is extremely sensitive to light -- so sensitive that when light is very intense, the receptors of the retina can be permanently damaged. A common example of this is the blindness that occurs subsequent to viewing the solar eclipse.

Ham and colleagues (1982) conducted experiments with rhesus monkeys (who had their lenses surgically removed) to determine which wavelengths of light are most damaging to the retina. Because rhesus monkeys have a retina that is nearly identical to that of humans, the results of these experiments can be generalized to humans. The results are presented in Figure 2.15 in terms of relative sensitivity to damage as a function of wavelength.

Damage observed by Ham et al. occurred to the receptors as well as to the cells behind the receptors that are necessary for receptor function, cells in the layer known as the retinal pigment epithelium. The data points in Figure 2.15 indicate that any wavelength of light, in sufficient intensity, can be damaging to the retina. Note, however, that the short wavelengths in the visible spectrum (ca. 450 nm) and the ultraviolet wavelengths (300 to 400 nm) are most

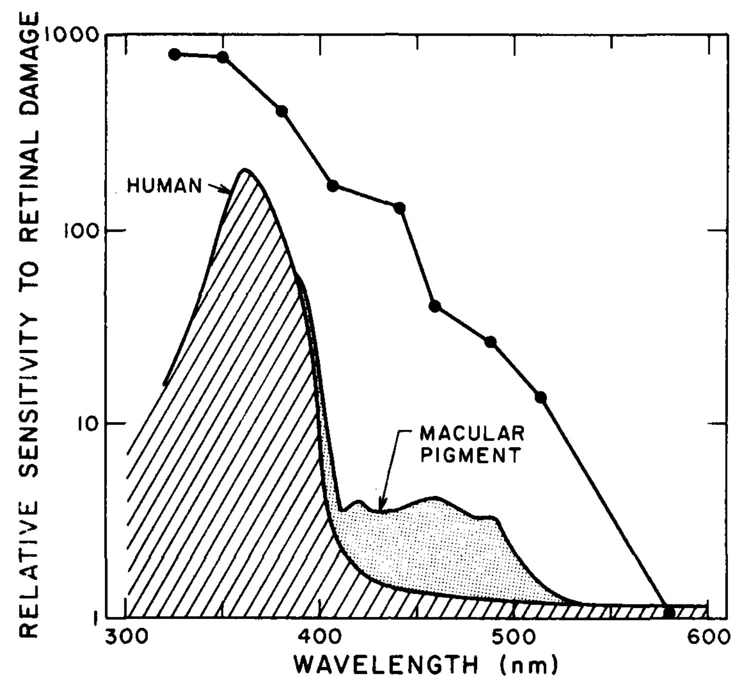


Figure 2.15. Relative sensitivity to retinal damage plotted as a function of wavelength. (data from Ham et al., 1982, original figure)

effective in producing damage.

The absorption of light at different wavelengths by the human lens and the macular pigment, a yellow pigment concentrated around the fovea, is indicated in Figure 2.15 by the hatched and screened areas, respectively. Since these pigments absorb the light indicated by the areas shown, they substantially reduce the intensity of the most hazardous wavelengths before that light reaches the retinal receptors. Thus, our lens and macular pigment provide a natural source of protection from light damage. Unfortunately, these natural filters do not always provide sufficient protection against the hazardous effects of ultraviolet radiation and many researchers advise additional protection against the long-term effects of radiation which may accumulate over the life span and contribute to aging of the retina (Werner, Peterzell & Scheetz, 1990) and possibly certain diseases of the retina such as age-related macular degeneration (Young, 1988). Because the intensity of ultraviolet radiation increases with increasing altitude, these concerns may be especially important to airline pilots.

Eye Movements

The field of view for humans is about 180° for the two eyes combined, as shown in Figure 2.16, but as we have seen in Figure 2.17, the receptor mosaic

varies with retinal eccentricity and so we rely on a smaller portion of the retina for processing detailed information. In particular, the act of fixation involves head and eye movements that position the image of objects of interest onto the fovea. This can be seen by recording the eye movements of someone who is viewing a picture. Figure 2.17 shows some recordings made by Yarbus (1967) of eye movements while viewing pictures. Notice that the eye moves to a point, fixates momentarily (producing a small dot on the record), and then jumps to

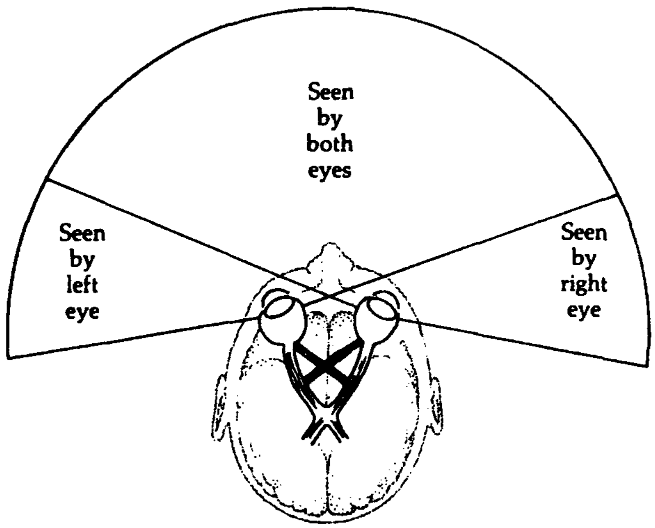


Figure 2.16. Visual field for humans: about 180°. (from Sekuler & Blake, 1985)

another point of interest. Notice also that much of the fixation occurs to features or in areas of light-dark change. Homogeneous areas normally do not evoke prolonged inspection. For information to be recognized or identified quickly and accurately, movements of the eye must be quick and accurate. This is accomplished by six muscles that are attached to the outside of each eye. These muscles are among the fastest in the human body.

There are two general classes of eye movements: vergence and conjunctive. Movements of the two eyes in different directions -- for example, when both eyes turn inward toward the nose -- are called *vergence* movements. These movements are essential for fixating objects that are close. The only way both eyes can have a near object focused on both foveas is by moving them inward. Eye movements that displace the two eyes together relative to the line of sight are known as ***conjunctive*** eye movements. There are three types of conjunctive eye movements: saccadic, pursuit, and vestibular. ***Saccadic*** eye movements are easily observed when asking a person to change fixation from one point in space to another. A fast ballistic movement is engaged to move the eye from

one point to the next. Careful measurements show that the delay between presentation of a peripheral stimulus and a saccade to that stimulus is on the order of 180 to 250 msec. The movement itself only requires about 100 msec for the eyes to travel a distance of 40° (Alpern, 1971). Saccadic movements of the eyes are necessary to extract information from our environment. For

Figure 2.17. Eye movements while viewing pictures; small dots are fixations. (from Yarbus, 1967)

example, during reading we may make as many as four small saccades across a line of type and one large saccade to return the eye to the beginning of the next line. We engage in many thousands of saccadic eye movements each day.

One of the great mysteries in eye movement research has to do with why we don't notice our eye movements. If the visual image in a motion picture were moved around the way the eyes move the visual image, it would be very disconcerting. The same motion of the image due to movement of the eye results in the appearance of a stable world. Part of the reason that saccadic eye movements are not disruptive has to do with an active suppression of visual sensitivity for about 50 msec before and after a saccadic eye movement (Volkmann, 1962). A similar reduction in visual sensitivity also occurs during blinks (Riggs, Volkmann & Moore, 1981), which is probably why we do not notice "the lights dimming" for one-third second, every four seconds, which is about the duration and frequency of eye blinking. The light is reduced by about 99% during a blink, but this change is seldom noticed.

Still another reason we may fail to notice blurring during a saccadic eye movement is due to ***visual masking***. When two stimuli are presented in quick succession, one stimulus may interfere with seeing the other. For example, threshold for detecting a weak visual stimulus will increase if a more intense stimulus is presented just before or just after the weak stimulus is presented. Similarly, the sharp images seen just before and after an eye movement may mask the blurred stimulus created during the saccade (Campbell & Wurtz, 1978).

While saccadic eye movements allow the eye to "jump" from one point to another, ***pursuit*** eye movements allow the eye to move slowly and steadily to fixate a moving object. These movements are very different from saccades and are controlled by different mechanisms in the brain. Saccadic eye movements are programmed to move the eye between two points with no changes in the direction of movement once the saccade has begun. Pursuit movements require brain mechanisms to determine the direction and velocity of a moving object for accurate tracking. Indeed, accurate tracking for slow moving objects is possible, but the accuracy decreases with increasing target speed.

Vestibular movements of the eye are responsible for maintaining fixation when the head or body moves. To maintain fixation during head movement, there must be compensatory changes in the eyes. The movement of the head is detected by a specialized sensory system called the vestibular system, and head-position information is relayed from the vestibular system to the brainstem areas controlling eye movements. Although we are seldom aware of vestibular eye movements, they are essential for normal visual perception. Some antibiotics

have been known to temporarily impair function of the vestibular system and eliminate vestibular eye movements. Under these conditions, it is virtually impossible to read signs or recognize objects because the lack of eye movements to compensate for head movements makes the world appear to jump about.

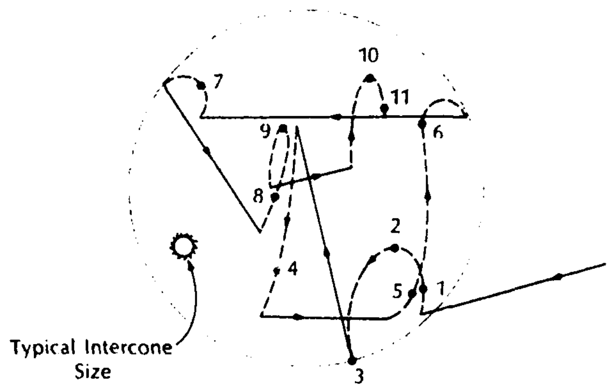


Figure 2.18. Eye movement records illustrating physiological nystagmus. (from Ditchburn, 1955)

Even when we are intently fixating an object, small random contractions of the eye muscles keep the eyes moving to some extent. These tiny eye movements are known as ***physiological nystagmus*** and include tiny drifting eye movements and microsaccades (Ditchburn, 1955). Physiological nystagmus is illustrated by eye movement recordings shown in Figure 2.18. The numbered dots represent successive time intervals of 200 msec. The large circle encompasses only 5 minutes of arc, so the movements are quite small, some on the order of the diameter of two photoreceptors.

One might wonder what would happen if the eyes did not move at all. To answer this question, Riggs et al. (1953) designed a clever apparatus in which the observer wore a contact lens with a mirror attached to it. Light from a projector that bounced off the mirror was projected onto the wall. Thus, when the eye moves, the mirror moves and, of course, so does the visual stimulus. As a consequence, the projected image always falls on the same part of the retina, and is called a ***stabilized retinal image***. The visual experience with a stabilized image is startling. Borders fade away and eventually the entire visual image disappears. In other words, when the retina is uniformly stimulated, the eye becomes temporarily blind to the image. Small movements of the eye destabilize the retinal image and make vision possible.

The fact that stabilized images disappear explains why we don't see the blood vessels in our own eyes. Figure 2.19, for example, shows the blood vessels in the eye which lie in front of the receptors. This means that when light passes into the eye and strikes the vessels, a shadow is cast on the retina. **Because this shadow moves wherever the eye moves, it is stabilized and we don't see it.** You can actually see the blood vessels in your eye by doing the following. Take a small flashlight and position it close to the outside corner of your eye. Look

straight ahead in a dark room and shine the light directly into the eye while moving it quickly back and forth. The moving light causes the shadows to move and hence the images are no longer stabilized and become visible.

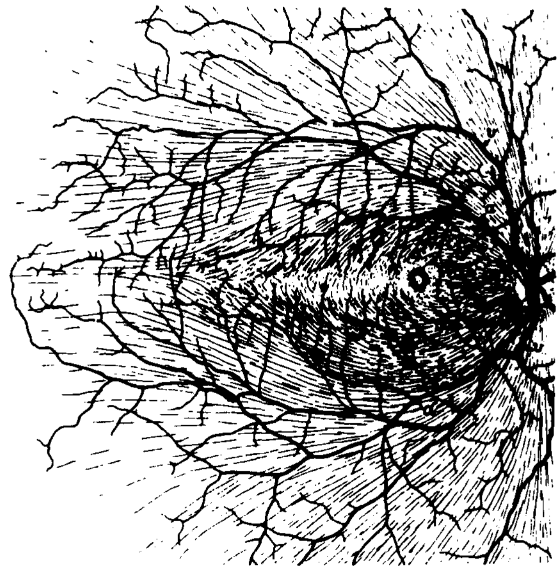


Figure 2.19. The dark lines show retinal blood vessels. The central circle, relatively free of blood vessels, represents the fovea. (from Polyak, 1941)

Temporal Vision

Many visual stimuli change over time, and the change itself can provide compelling information about the stimulus. Indeed, sometimes it is only the temporal variation of a stimulus that allows it to be detected, discriminated, or recognized.

Flicker

If a light is turned on and off in rapid succession, we will experience a sensation that we call flicker. If the frequency of oscillations, measured in cycles per second (cps or Hz), is high enough, the flicker will no longer be perceptible. This is known as the ***critical flicker fusion (CFF)*** frequency. At high light levels, CFF may occur at frequencies as high as 60 Hz. The fact that flicker fuses at high frequencies explains why fluorescent lamps appear to be steady even though they are going on and off at 120 cps. Here, we will discuss some of the main parameters that determine CFF; see Brown (1965) for a complete review of the literature.

Our ability to detect flicker depends on the light level; as luminance increases, flicker is easier to detect. Figure 2.20 shows CFF as a function of luminance (Hecht & Smith, 1936). It is clear that CFF depends on both the light level and the stimulus area. When the area is large enough to stimulate both rods and cones, the curve has two branches. When cones dominate sensitivity, CFF increases linearly with light level over a wide range before reaching an asymptote. The lower branch of each two-part curve is mediated by rods which have relatively low sensitivity to flicker.

The data shown in Figure 2.20 were obtained by having the subjects view the center of the stimulus. The effect of increasing area was, therefore, partially confounded with retinal location. You may have noticed this yourself under nonlaboratory conditions; when looking directly at a large object like a

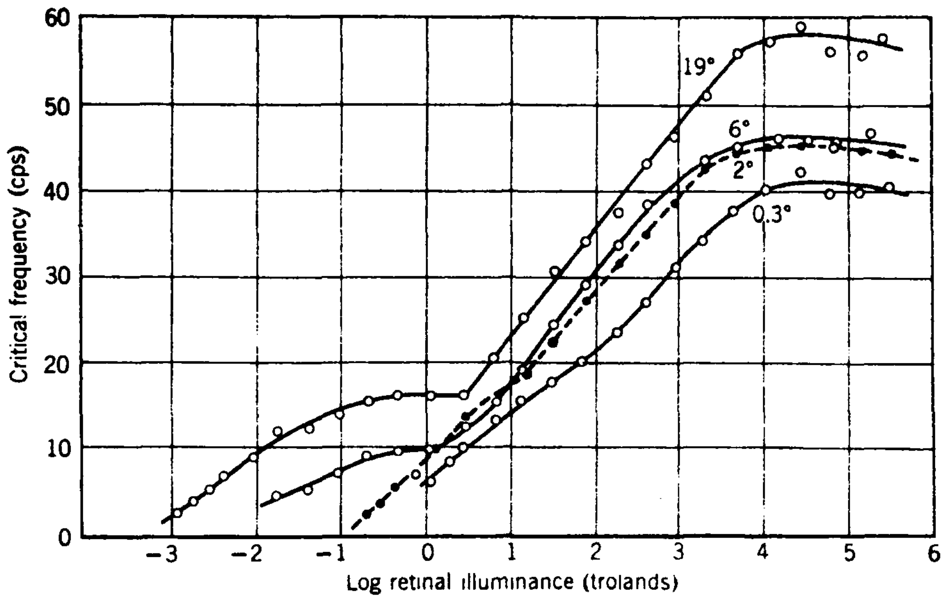


Figure 2.20. Critical flicker fusion for a centrally viewed stimulus plotted as a function of log luminance. Different curves show different stimulus sizes. (from Hecht & Smith, 1936)

computer screen it may appear steady with direct viewing, but not in your periphery. A flickering stimulus (e.g., part of a display) in the periphery can be very distracting as it efficiently attracts attention.

In Figure 2.21, data are presented showing how sensitivity to flicker varies with the intensity of the stimulus and retinal location (Hecht & Verrijp, 1933). Data were obtained with a 2° stimulus that was viewed foveally (0° in the figure) and at 5° and 15° eccentric to the fovea. It appears that a single curve can account for the changes in CFF with intensity in the fovea, but to describe CFF at more peripheral locations requires curves with two branches. From the data in the figure, one can see that the relationship between CFF and retinal location is complex. At high light levels there is a decrease in CFF from the fovea to the periphery, whereas the reverse is true at low light levels. Flicker sensitivity declines rather markedly as a function of increasing observer age, as shown in Figure 2.22. This is to be expected at least in part because the light transmitted by the lens decreases with age, and flicker sensitivity is dependent on light level. It is still not entirely clear whether there are additional neural changes associated with age-related changes in CFF or whether these changes in flicker sensitivity are secondary to changes in light level alone (Weale, 1982). In any case, this means that a display may appear to be flickering for one observer while an older observer would see the display as steady, i.e., no flicker.

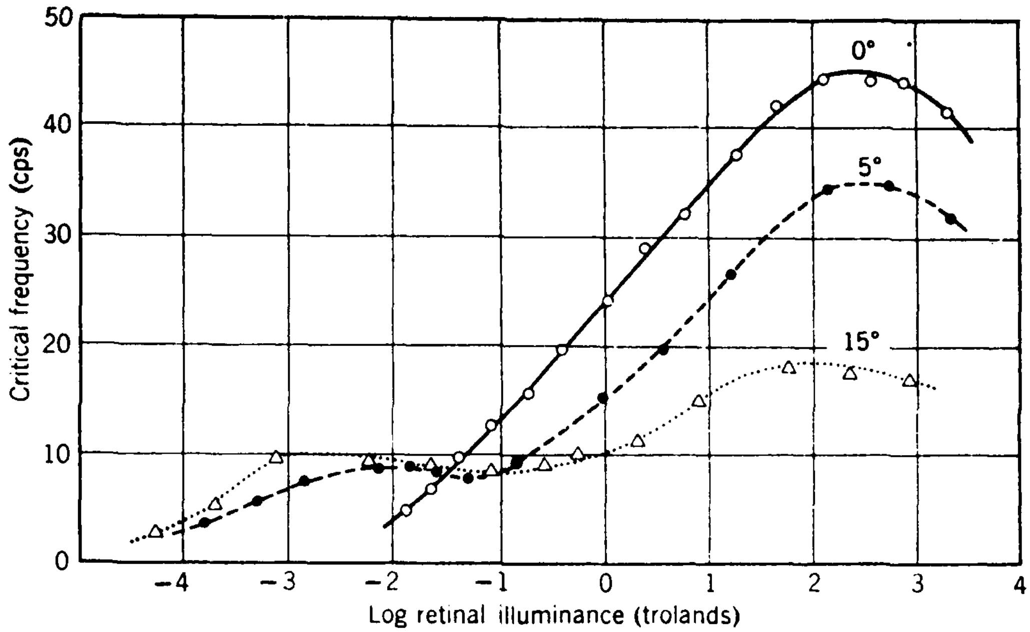


Figure 2.21. **Critical flicker fusion for a 2° stimulus plotted as a function of log luminance. Different curves show CFF for different retinal loci. (from Hecht & Verrijp, 1933)**

The CFF measurements discussed so far were obtained with a stimulus that was either completely on or completely off. If we were to draw a graph of the intensity over time it would look like shape 1 illustrated in Figure 2.23, and is known as a square wave. With specialized equipment, deLange (1958) also measured flicker sensitivity using other waveforms (changes in light intensity over time) that are shown by the inset in Figure 2.23, and at three different light levels. In each case, the stimulus was repeatedly made brighter and dimmer at the frequency specified on the horizontal axis. The vertical axis plots the "ripple ratio," or amplitude of modulation, which refers to the amount that the light must be increased and decreased relative to the average light level to *just detect flicker. Figure 2.23 thus illustrates our sensitivity to flicker at all* different frequencies. It can be seen that we are most sensitive to flicker at about 10 Hz. At higher frequencies, the amplitude of modulation must be increased in order for flicker to be detected.

We noted in our discussion of hearing that the response of the human auditory system to complex sounds can be predicted by decomposing the complex tones into a set of pure tones, or sinusoidal waveforms. deLange (1958) applied this approach to the different waveforms of his flickering stimuli by mathematically analyzing them in terms of a set of sine-wave components (using Fourier analysis). Figure 2.23 shows a plot of the amplitude of modulation for the

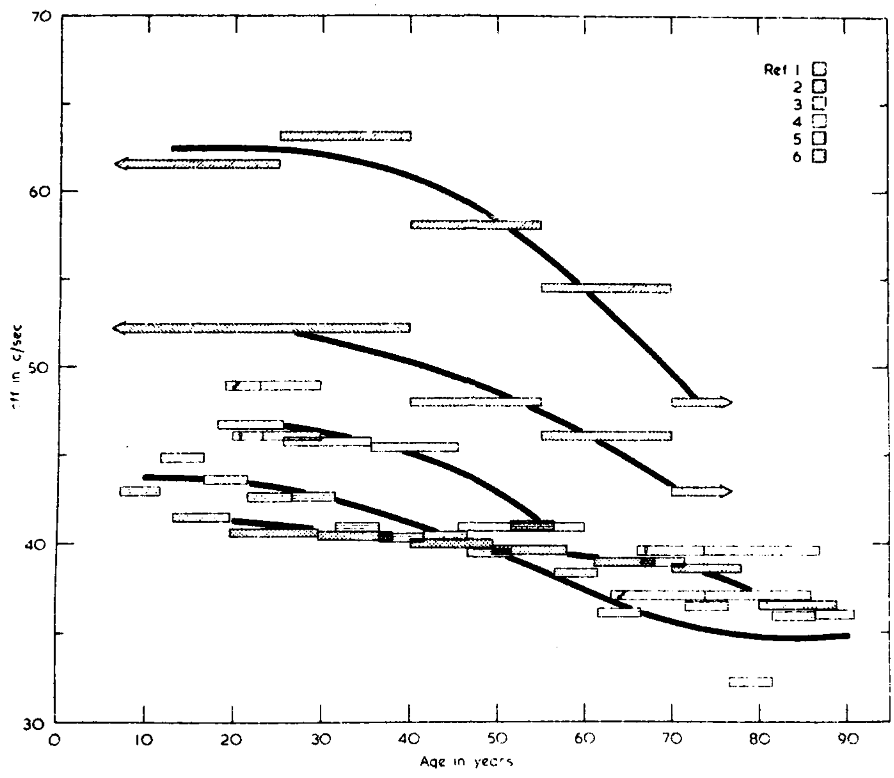


Figure 2.22. **Critical flicker fusion plotted as a function of age for six different studies which used different stimulus conditions. (from Weale, 1982)**

fundamental component of the different waveforms, i.e., the amplitude of the lowest frequency contained in the complex wave. When analyzed in this way, it appears as though sensitivity to flicker for complex waves can be predicted by the response of the eye to the various sinusoidal components into which the complex wave can be decomposed.

Motion

If you doubt that motion is a fundamental perceptual quality, try to imagine what life would be like without the ability to experience it. A rare case of damage to a part of the brain (called area MT) that appears to be specialized for analysis of motion occurred in a woman in Munich. The scientists who studied this woman noted what it was like:

> She had difficulty, for example, in pouring tea or coffee into a cup because the fluid appeared to be frozen, like a glacier. In addition, she could not stop pouring at the right time since she was unable to perceive the movement in the cup (or a pot) when

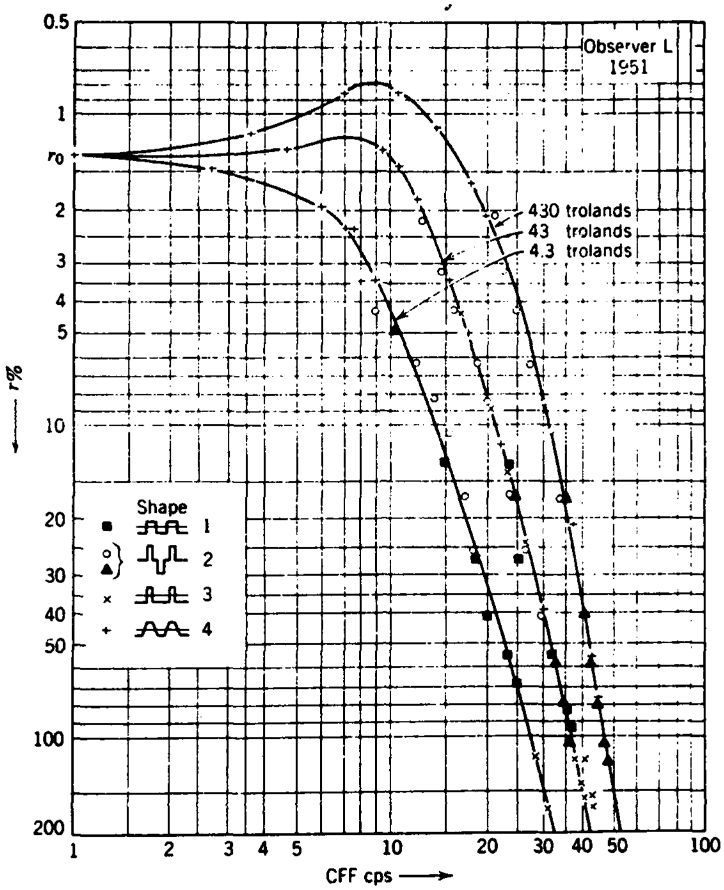
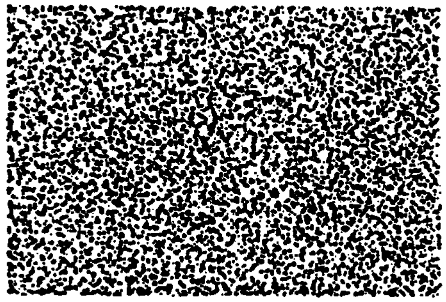


Figure 2.23. Modulation amplitude (r%) of the fundamental component contained in the waves plotted as a function of flicker frequency. (from deLange, 1958)

> the fluid rose.... In a room where more than two other people were walking she felt very insecure and unwell, and usually left the room immediately, because 'people were suddenly here or there but I have not seen them moving.'... She could not cross the street because of her inability to judge the speed of a car, but she could identify the car itself without difficulty. 'When I'm looking at the car first, it seems far away. But then, when I want to cross the road, suddenly the car is very near.'
> (From Zihl, von Cramon & Mai, 1983, p. 315).

Figure 2.24 shows a square comprised of dots that are arranged in a random order, and a set of dots arranged so that they spell a word. If the two sets of dots are printed on a transparent sheet and superimposed, no word can be read and one observes only a set of dots. However, if one sheet moves relative to the other, the dots that move together form a clearly legible word. Structure



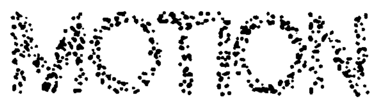


Figure 2.24. If the two sets of dots are superimposed, no pattern can be detected. However, if one set of dots moves relative to the other, the word "motion" will be clearly visible.

emerges from the motion information. This illustrates one of the many functions of motion -- to separate figure and ground. When an object moves relative to a background, the visual system separates the scene into figure and ground.

Our perception of motion is influenced by many factors. Our perception of motion speed is affected by the sizes of moving objects and background. Measures of motion thresholds indicate that we can detect changes of an object on a stationary background on the order of 1 to 2 minutes of arc per second. However, when the background cues are removed, motion thresholds increase by about a factor of ten (see Graham, 1965b). These thresholds also depend on the size of the moving object and background. For example, Brown (1931) compared movement of circles inside rectangles of different size, as illustrated by Figure 2.25. Observers were asked to adjust the speed of one of the dots to match the experimenter-controlled speed of the other. He found that in the large rectangle, the spot had to move much faster than in the small rectangle to be perceived as moving at the same speed. As a general rule, when different size objects are moving at the same speed, the larger one will appear to be moving more slowly than the small one. Leibowitz (1983) believes that this is the reason for the large number of fatalities at railroad crossings. Large locomotives are easily seen from the road, but they are perceived to be moving more slowly than they really are. As a consequence, motorists misjudge the amount of time they have to cross the tracks.

Most of the motion that we observe involves actual displacement of objects over time, but this is not a necessary condition for the experience of motion. For example, a compelling sense of motion occurs if we view two lights, separated in space, that alternately flash on and off with a brief time interval between the flashes (about 60 msec). This movement is called ***stroboscopic motion***, and it is

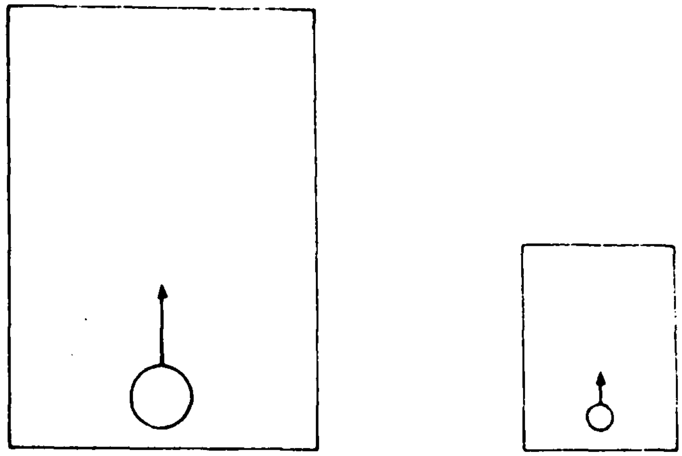


Figure 2.25. Illustration of experiment by Brown (1931). The left circle must move faster than the one on the right for the two to be perceived as moving at the same speed. (from Goldstein, 1984)

very important for motion pictures because films are merely a set of still pictures flashed in quick succession. Stroboscopic movement is also important for understanding how motion is actually perceived because it demonstrates that it is a perceptual quality of its own, rather than a derivative of our sense of time and space.

In early studies of stroboscopic movement, Wertheimer (1912) discovered that the apparent movement of two spots of light in the above demonstration goes through several different stages depending on the time interval between the flashes. If the interval was less than 30 msec, no movement was detected. Between about 30 and 60 msec there was partial or jerky movement, while at about 60 msec intervals the movement appeared smooth and continuous. Between about 60 and 200 msec, movement could be perceived, but the form of the object could not (objectless movement). Above about 200 msec, no movement was detected. Of course, these values depend on the distance between the two stimuli, but at all distances the different stages could be identified.

Still another type of movement perception occurs without actual movement of the object. For example, ***induced movement*** occurs when a background moves in the presence of a stationary object, but it is the object not the background that is seen as moving. You may have had this experience looking at the moon when clouds were moving quickly in the wind; it is not unusual to have the experience of the moon moving across the sky.

On a clear and quiet night looking at a star against a dark sky you may also have experienced illusory movement of the star. The effect is easily demonstrated by looking at a small light on a dark background. It may start to move, even though it is rigidly fixed in place. This illusory movement is known as the ***autokinetic effect***. It is not well understood, but some researchers believe it may be due to drifting movements of the eyes (Matin & MacKinnon, 1964). Whatever the cause, one can imagine practical situations in which the autokinetic effect has the potential to cause errors in judgment.

Chapter 3

Color Vision

by John S. Werner, Ph.D., University of Colorado at Boulder

Color Mixture

From the scotopic spectral sensitivity curve (Figure 2.12, p. 22) it is clear that rods are not equally sensitive to all wavelengths. Why, then, do all wavelengths look the same to us when they stimulate only the rods? The answer is that a rod can only produce one type of signal regardless of the wavelength that stimulates it. That is, all absorbed quanta have the same effect on a single receptor, and, therefore, it can only pass on one type of signal to the brain. Thus, even though some wavelengths are more easily absorbed than others, once absorbed they all have the same effect.

If each receptor cell has only one type of response, what explains how we use our cones to see color? The answer is that we have three different types of cones. They differ because each type contains a different photopigment. Figure 3.1 shows the absorption spectra -- plots of relative absorption as a function of wavelength -- for the three types of photopigment contained in human cones. Note that each type is capable of absorbing over a broad wavelength range. One type maximally absorbs quanta at about 440 nm, another at about 530 nm, and the third type at about 560 nm. We call these three types of receptors short-, middle-, and long-wave cones, based on their wavelength of maximal sensitivity.

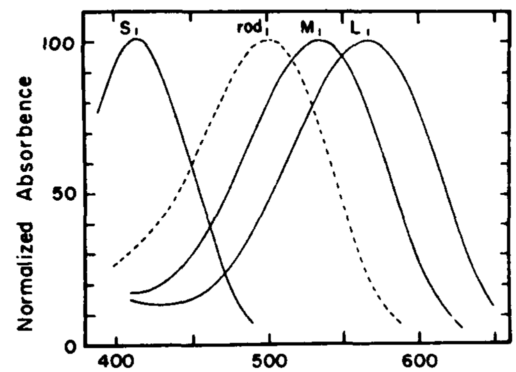


Figure 3.1. Absorption of the cone and rod photopigments plotted as a function of wavelength. The curves have been normalized to the same heights. (after Bowmaker et al., 1980)

Now suppose we look at two monochromatic lights presented side-by-side. If the wavelengths are 450 and 605 nm respectively, we would probably describe the lights as reddish blue and yellowish red. Note that these two wavelengths are equally absorbed by the middle-wave cones. The same quantal absorption for two lights means that a single receptor must produce the same signal for the two lights. The 450 nm light will, however, elicit a much stronger signal in the short-wave cones than in the long-wave cones, and the opposite will occur for the 605 nm light. Thus, both monochromatic lights will produce signals in all three cone types, but the pattern of activity will differ among them. This pattern of receptor activity is transmitted to the brain and allows us to discriminate a difference in the two wavelengths.

Because our cone system produces these different patterns of response to different wavelengths, it can distinguish changes in intensity and wavelength. But not all differences can be discriminated. In the mid-1800s, Helmholtz and Maxwell performed experiments by having a subject match a light composed of three different wavelengths with a light containing only one wavelength. They discovered that any single wavelength of light can be perfectly matched by a mixture of three other wavelengths. This match is possible because the three combined wavelengths produce the same pattern of activity in the different cone types that is produced by the one wavelength alone. Thus, an observer perceives the two physically different patches of light as identical.

Our three types of cone receptors allow us to discriminate different wavelengths, but the example above showed how this system can be fooled. Actually, it is this very limitation that allows us to have electronic color displays. The image on the display consists of many small spots of light, or pixels. Three contiguous pixels, containing different phosphors, may produce either red, green, or blue. These three phosphors are so small and close together that the light produced by them is blended in the retinal image. The colors on the display are created by electrically exciting these phosphors to produce the amounts of the three lights that produce the color we see.

Variation in Cone Types with Retinal Eccentricity

Figure 2.10 (Chapter 2, p. 20) showed the distribution of rods and cones with varying eccentricity. A careful examination of the distribution of cones would show that there are asymmetries in the distribution of cones. At any given eccentricity, the nasal retina has a higher density of cones than the temporal retina. There appear to be no asymmetries along the superior to inferior meridian. The practical consequences of the retinal asymmetry in cone distribution are not clear, although it has been shown that color vision is, in some sense, better in the nasal compared to the temporal retina (Uchikawa, Kaiser & Uchikawa, 1982).

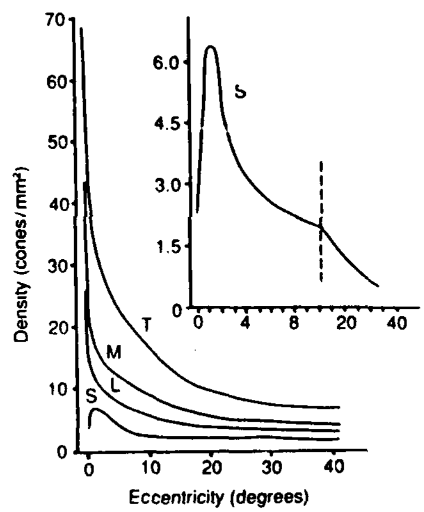


Figure 3.2 The number of short-, middle-, and long-wave sensitive cones per square mm in a baboon retina as a function of retinal eccentricity. (after Marc & Sperling, 1977)

The distribution of the three cone types also varies with retinal eccentricity, as shown by Figure 3.2. The data presented in this figure are actually from a baboon retina and are believed to be similar to the human cone distribution with an important exception. Whereas the baboon has more M than L cones, humans have more L than M cones. In fact, for the central 2° of retina, the ratio of L:M:S cones is about 32:16:1 (Vos & Walraven, 1971). The relative scarcity of S cones has important implications for visual perception. Partly because of their numbers and partly because of their neural connections, the S cones make a negligible contribution to high spatial acuity and high temporal sensitivity (Kelly, 1974). The S cones are important for color discriminations.

The inset of Figure 3.2 shows a magnified scale of the retinal distribution of S cones. There are virtually no S cones in the center of the fovea. This means

that color discriminations that depend on S cones will be impaired if the image is sufficiently small to fall only on the center of the fovea. This is illustrated by Figure 3.3. When viewed close, so that the visual angle of each circle subtends several degrees, it is easy for an individual with normal color vision to discriminate the various pairs; yellow vs. white, blue vs. green, and red vs. green. Viewed from a distance of several feet, however, the yellow and white, as well as the blue and green, pairs will be indiscriminable. This is called ***small-field tritanopia***, because tritanopes are individuals who completely lack S cones. A tritanope would not be able to discriminate the yellow from the white in Figure 3.3 regardless of their sizes. With certain small fields, even normal individuals behave like tritanopes. Notice that with the small field condition, the red-green pair is still discriminable because S cones are not necessary for this discrimination. Thus, the small-field effect is limited to discriminations that depend on S cones.

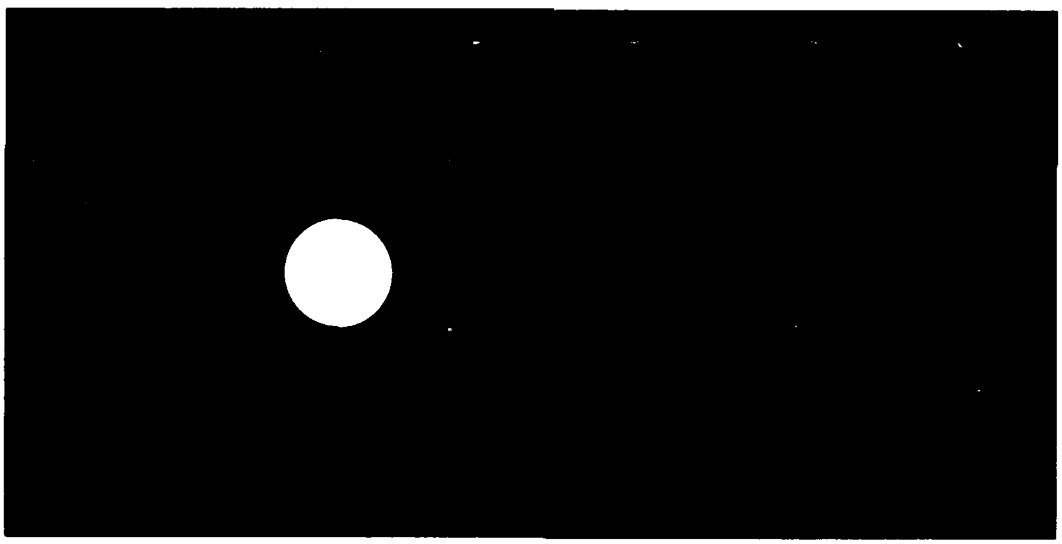


Figure 3.3. **Colors (yellow and white) not discriminable at a distance due to small field tritanopia.**

No blue and green pairs are shown on this page. There is no reference to blue and green pairs in the new text. Refer to the enclosed errata sheet for the correct text.

Color Vision Deficiencies

We take the colorfulness of our world so much for granted that it is hard to imagine a form of color vision different from our own. Normal color vision is based on three types of cone receptors, and such individuals are known as ***trichromats***. An individual can be classified as trichromatic if he or she requires a mixture of three lights (known as primaries) to match all wavelengths of the spectrum.

Congenital Deficiencies

Many individuals require three primaries to match all wavelengths of the spectrum, but the intensity ratio of the three lights needed is not normal. Such individuals are called ***anomalous trichromats***. The reason for anomalous trichromacy is that one or more of the cone receptor classes contains a photopigment that is shifted along the wavelength scale relative to normals. Since there are three types of cones, there can be at least three types of anomalous trichromacy, depending on which type of photopigment is shifted.

Anomalous trichromats can be classified as ***tritanomalous*** (shifted pigment in the short-wave cones), ***deuteranomalous*** (shifted pigment in the middle-wave cones) or ***protanomalous*** (shifted pigment in the long-wave cones). Tritanomalous color vision is extremely rare -- so rare that some authorities doubt its existence. Deuteranomaly and protanomaly are not rare, as can be seen in Table 3.1. In both of these forms of anomalous vision, the middle- and long-wave pigments overlap in their sensitivity by a greater degree than normal. This affects not only their color matching, but also the ability of anomalous trichromats to discriminate certain wavelengths of light. A more severe form of color deficiency exists when an individual is completely missing one type of photopigment in the cones. It should be mentioned that the normal number of cones is present in such individuals, but the cones are segregated into two classes rather than three. These individuals are called ***dichromats*** because they require only two primaries to match all wavelengths of the spectrum. There are three types of dichromat. A person who is missing the short-wave cone photopigment is called a ***tritanope***, and would have difficulty discriminating white from yellow, for example. Persons missing the normal middle-wave cone photopigments are known as ***deuteranopes*** and would not be able to discriminate red from green based on wavelength alone (see Figure 3.3). Red-green discriminations are also impaired in ***protanopes***, individuals missing the normal long-wave cone photopigment. Finally, there are some individuals, known as ***monochromats***, who require only one wavelength of light to match all others of the spectrum. This implies that the individual is using only one type of receptor in color matching. Such a person could be a monochromat due to having only one type

Table 3.1
Congenital Color Vision Deficiencies

Type	Abnormality	Prevalence
Tritanomaly	Shifted S-Cone Pigment	Males: ? Females: ?
Deuteranomaly	Shifted M-Cone Pigment	Males: 5.1% Females: 0.5%
Protanomaly	Shifted L-Cone Pigment	Males: 1.0% Females: 0.02%
Tritanopia	Missing S-Cone Pigment	Males: 0.0007% Females: 0.0007%
Deuteranopia	Missing M-Cone Pigment	Males: 1.1% Females: 0.01%
Protanopia	Missing L-Cone Pigment	Males: 1.0% Females: 0.02%

S-Cone: short-wave cone M-Cone: middle-wave cone L-Cone: long-wave cone

of cone (a cone monochromat) or because the individual has no cones (a rod monochromat). The cone monochromat has one type of cone for photopic vision and rods for scotopic vision. The rod monochromat has no cones so is severely impaired in functioning under photopic (day vision) conditions.

When anomalous trichromacy or dichromacy is present from birth, the deficiency is called congenital. The incidence of all forms of color vision deficiency combined varies across populations; about 8% in Caucasian males, 5% among Asian males, and only 3% in Black and Native American males. Table 3.1 summarizes the incidence of congenital color vision deficiency in North America and Western Europe. It is clear that these forms are inherited with the most common forms carried by the sex chromosomes. This is why the incidence of middle- and long-wave cone deficiencies is about ten times more prevalent in males than in females.

Acquired Deficiencies

Not all deficiencies of color vision are congenital, some are acquired in later life. Unlike congenital deficiencies which are due to abnormalities at the level of the photopigments, acquired deficiencies can be due to disruption of processing at any level of the visual system. For example, on rare occasions following a stroke, an individual may experience damage to a particular region of the brain involved in color processing that will render him or her permanently color blind. Such a case was reported for a customs official who

had passed color vision screening tests as a condition of employment, but he could not do so after his stroke (Pearlman, Birch & Meadows, 1971). The man had good memory for colors, but when given crayons to color a picture, he appeared completely confused in his selections. Fortunately, such cases of cortical color blindness are extremely rare.

Other acquired deficiencies of color vision are not rare. Glaucoma and diabetes, for example, often impair functioning of S cones (Adams et al., 1987). In some cases, these changes in color vision occur before there are any physical changes that can be detected by standard clinical testing and before there are changes in visual acuity. Many acquired defects of color vision do not fit neatly into the categories of color deficiency that are used to classify congenital losses (Verriest, 1963). Early in the disease a loss of yellow-blue discrimination is typically noticed, but this may be followed by impairment of red-green discriminations. The incidence of acquired defects of color vision in the population has been estimated at about 5%, but these figures are not unequivocal.

Some drugs (both recreational and prescription) can cause defects of color vision. For example, blue-yellow color defects have been associated with certain medications used in the treatment of psychiatric disorders (e.g., phenothiazine [Thorazine] and thioridazine hydrochloride [Mellaril]). These effects can persist even after the medication is withdrawn. A more commonly used drug, chloroquine, prescribed as an antimalarial drug, has also been associated with blue-yellow defects. Red-green defects have also been reported as a side effect of certain medications. Among the drugs involved are certain antibiotics such as streptomycin and cardiovascular drugs such as Digoxin. The list of drugs that may impair color vision is actually quite large (see Pokorny et al., 1979), but the patient is seldom made aware of this possible side effect.

Variation with Age

We have seen that as we get older, the lens of the eye becomes less efficient at transmitting light, particularly light at short wavelengths. Since color discrimination is impaired by a reduction in light intensity, it is perhaps not surprising that performance on color vision tests can change with age. Verriest (1963) has shown age-related losses in performance on the Farnsworth-Munsell 100-Hue test of color discrimination, and these changes can be mimicked with young observers who are tested with short-wave absorbing filters placed in front of their eyes. In general, these losses in discrimination of the elderly are similar to deficits associated with congenital deficiencies of short-wave cones.

While it might be thought that changes in color vision with advancing age are only secondary to reductions in light transmission by the lens, this is not the case.

Werner and Steele (1988) measured the sensitivity of each of the three classes of cones using subjects between the ages of 10 and 84 years. Figure 3.4 presents a summary of their results. Each symbol represents a different observer's cone sensitivity, and each panel represents one of the three types of cone receptors. You can see that there are large individual differences at each age, but there is also a significant reduction in sensitivity throughout life. Converting from the logarithmic scale used to plot the data, there is a reduction of approximately 25% in cone sensitivity for each decade of life. This means that our ability to distinguish between subtly different colors deteriorates with age.

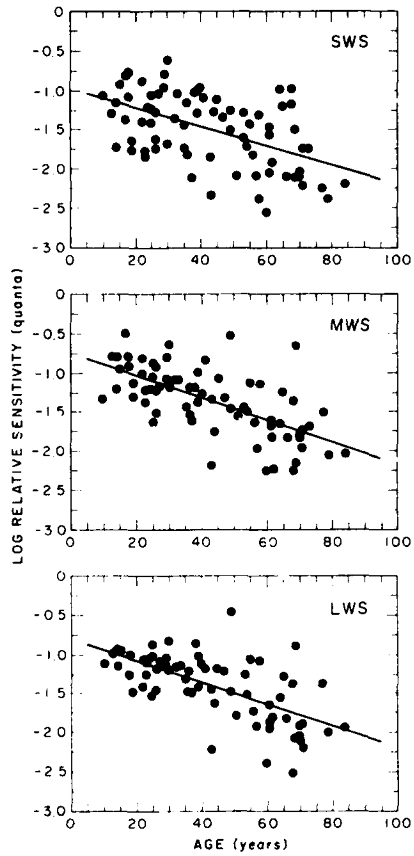


Figure 3.4. Log sensitivity of short-, middle-, and long-wave cones, measured psychophysically, plotted as a function of observer age. (data from Werner & Steele, 1988, figure from Werner et al., 1990)

Testing

Individuals with abnormal color vision are often unaware that their color vision differs from normals. Even dichromats can often name colors quite well in their natural environment because reds and greens, or blues and yellows, for example, may differ in their brightness. Thus, to properly test for color vision deficiencies, special tests are required.

The most definitive way to measure color vision is through color matching. A yellow light (590 nm) can be matched with a mixture of a yellowish red (670 nm) and a yellowish green (545 nm). The stimulus used for such a test is illustrated by Figure 3.5 and is produced by an instrument called an anomaloscope. Deuteranomalous and protanomalous individuals will differ from normal in the ratio of the two light intensities in the mixture that is required to match the yellow. Deuteranopes and protanopes can match the yellow using only one of the two lights simply by adjusting the intensity. Other wavelength mixtures can be used to diagnose deficiencies of the short-wave cones.

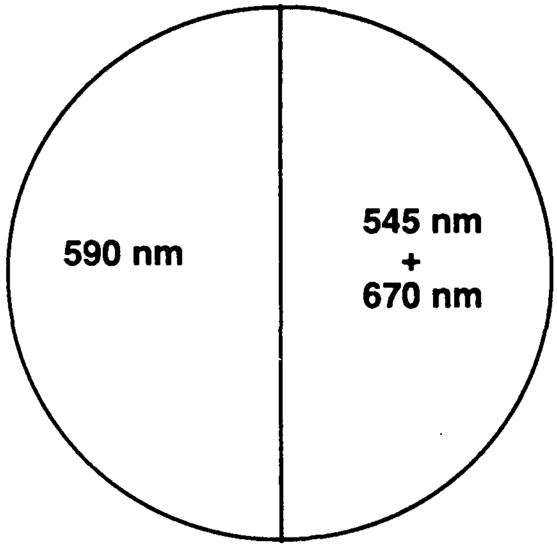


Figure 3.5. A schematic of the split field produced by an anomaloscope

Unfortunately, anomaloscopes are expensive and not readily available to most clinicians.

Perhaps the most familiar test for assessing color vision deficiency involves a series of plates composed of dots of one color which form a number or simple geometric form such as a circle or square. Surrounding these dots are others of a different color. The dots are carefully chosen so that, when illuminated with the proper lamp, normal individuals will be able to see the number or form but individuals with color vision deficiencies will not. Various color combinations are provided by different plates in order to detect different forms of deficiency.

Figure 3.6 shows one of these *pseudoisochromatic plates* used for testing color vision. Normal trichromats see a number 46 in this plate, but monochromats, certain dichromats and anomalous trichromats will not. This test provides an assessment of deficiencies involving middle- and long-wave cones, but most of the plate tests are not useful for detecting deficiencies of short-wave cones. This means that individuals who confuse certain reds and greens are more likely to be identified than individuals who confuse yellows and blues (or yellows and whites).

To detect abnormalities of any of the three cone types, a clinician could use the Panel D-15 test shown in Figure 3.7. This test consists of a number of caps of different colors. The object of the test is to arrange the caps in a logical color sequence. One of the caps is fixed in the tray and the subject is asked to place the one that is most similar next to it in the tray, and then to place the next most similar near the second cap and so forth. Each of the different types of color deficient observers will choose a different arrangement of the caps, which

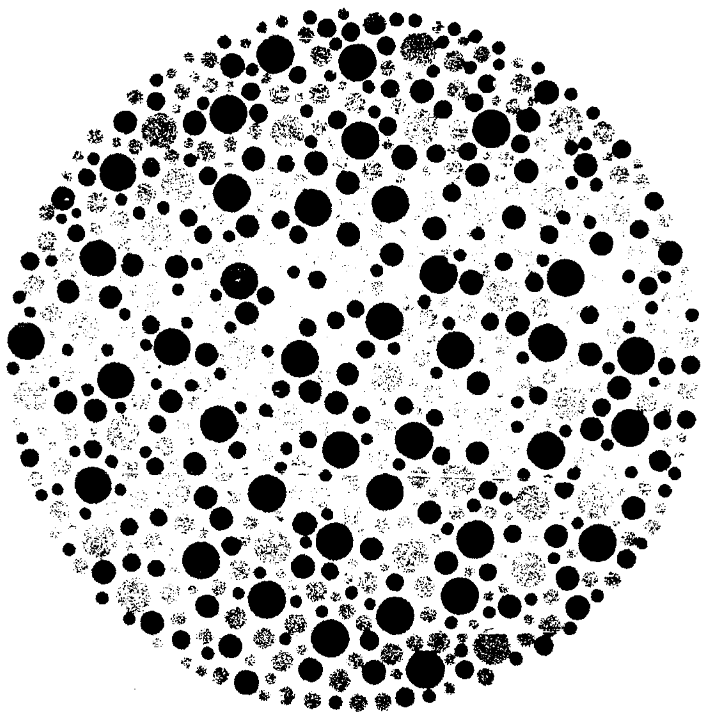


Figure 3.6. A pseudoisochromatic plate from the Dvorine Plate Test for color vision deficiencies.

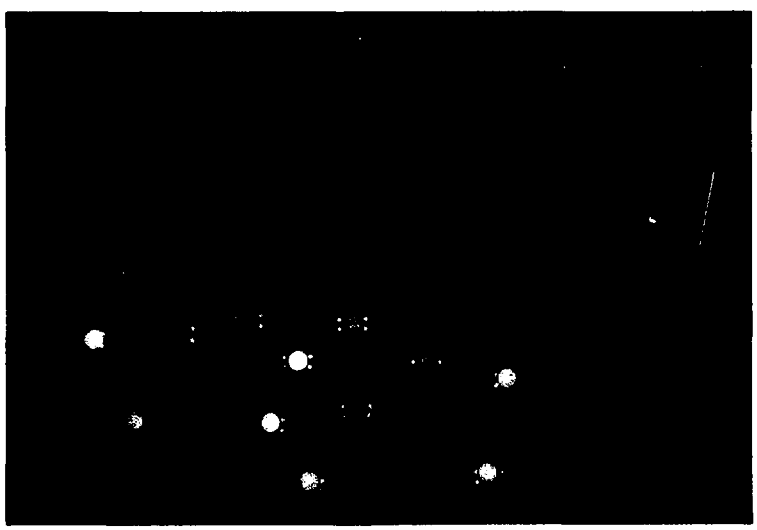


Figure 3.7. Farnsworth Dichotomous Test of Color Blindness, Panel D-15. Copyright © 1947 by The Psychological Corporation. Reproduced by permission. All rights reserved.

can be scored by reference to numbers on the bottom of the caps.

Note of Caution:

According to the Society of Automotive Engineers, ARP4032 (1988): "Approximately 3% of private pilots, 2% of commercial pilots, and 1% of airline transport pilots are known to have some form of color vision deficiency" (page 12). As already mentioned, individuals with abnormal color vision are often good at naming colors. People with such deficiencies learn to use other cues to discriminate colors; they learn, for example, that on a stop light, red is on top. Many color deficient observers could name the colors in most aircraft cockpits without having learned position cues. This does not, however, imply that they can process the colors normally. Discriminating between the colors may not be normal, especially under conditions in which the colors are desaturated ("washed out"). Search and reaction times are also impaired in color deficient observers. Cole and Macdonald (1988) demonstrated this using cockpit displays with redundant color coding (the meaning of the display symbols are coded by color and another cue such as shape).

Finally, we have already noted that screening for color vision deficiency requires certain tests, but it should be emphasized that these tests are only valid when administered under the proper conditions. The proper illumination of the tests can be obtained with specialized lamps, but because of their expense they are not always used. Failure to use the proper illuminant may result in misdiagnosis or failure to detect a color deficiency. Many of these testing considerations are summarized in a review by the Vision Committee of the National Research Council (1981).

Color Appearance

Color is defined by three properties: brightness, hue, and saturation. It would be convenient for engineers if these three psychological properties were related in one-to-one correspondence to physical properties of light, but they are not.

Imagine that you are sitting in a dark room viewing a moderately bright monochromatic light of 550 nm. A normal trichromat would say it is yellowish green. If we increased the number of quanta the light emits, you would say that the light is now brighter. What you experience as ***brightness*** increases with the light intensity, but before you conclude that brightness depends only on light intensity, look at Figure 3.8 which demonstrates ***simultaneous brightness contrast***. The two central patches are identical, but their brightness is influenced by the surroundings. All things being equal, brightness increases with intensity, but it is also affected by other factors.

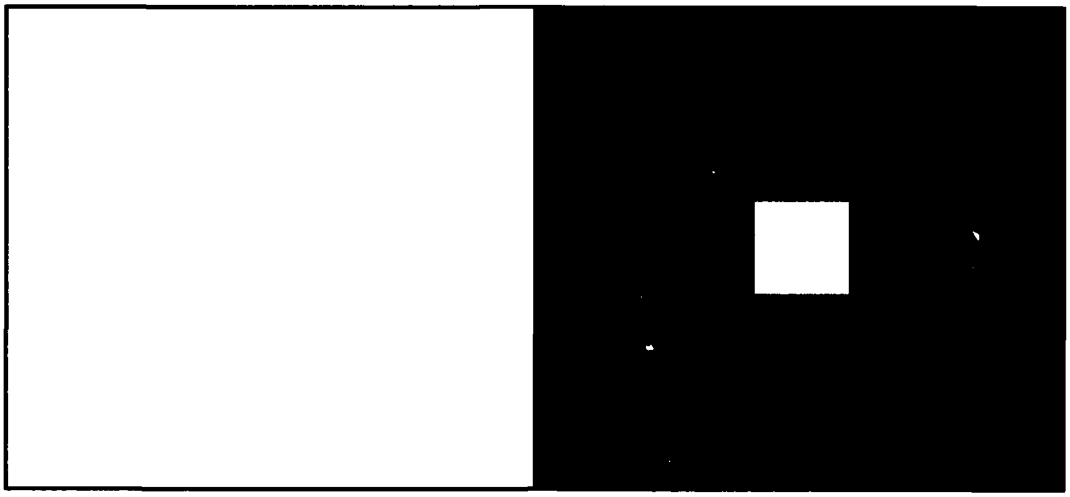


Figure 3.8. **An illustration of simultaneous brightness contrast.**

As we increase the intensity of our 550 nm light, you will probably notice that what appeared as green with just a tinge of yellow now has a much more vivid yellow component. You might say that the color has changed, but this change in appearance is described more precisely as a change in hue. ***Hue*** refers to our chromatic experience with light, such as redness and greenness. Many people think that particular wavelengths produce definite hues, but this is not entirely correct. Wavelength is related to hue, but one must consider other variables as well, such as intensity. In our example, a single wavelength produced somewhat different hues at different intensities.

A third change in the appearance of our 550 nm light as we increase the intensity is that the tinge of whiteness that was detectable at low intensities has now become clearer. The whiteness or blackness component is another dimension of our color experience known as ***saturation***. A light with little white is said to be highly saturated and appears vivid; a light with more whiteness is less saturated and appears more "washed out."

Thus, there are three dimensions of color experience: brightness, hue, and saturation. These dimensions are not uniquely related to quanta and wavelength. As we increased the number of quanta yet kept the wavelength constant, we saw a clear change in brightness, but also a change in hue and saturation.

Chromatic and Achromatic Colors

Suppose we look at two physically different spots of light that perfectly match, and that we call orange. The existence of three types of cone receptors in the color-normal person explains why the two colors cannot be discriminated, but it does not explain why we see the particular hue as orange. Hering (1920) proposed a theory to explain the appearance of hues. He proposed that all our experiences of hue can be reduced to four fundamental sensations: red, green, yellow, and blue. Thus orange is nothing more than a yellow-red. Consistent with this observation, modern experimental evidence has shown that the four basic terms are both necessary and sufficient to describe all hues. Figure 3.9 shows how these hue names are used to describe monochromatic lights from 400 to 700 nm. Notice that the percentage of red or green is plotted from 0 to

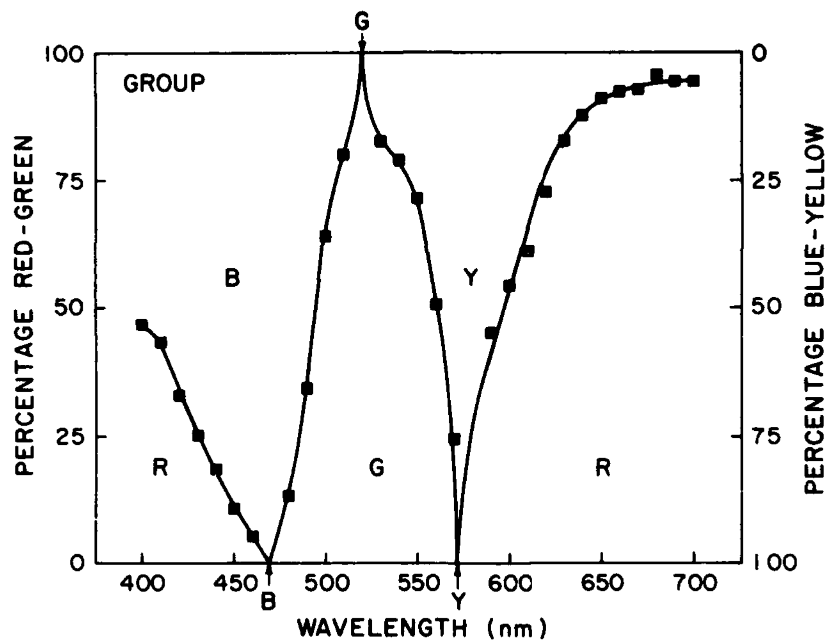


Figure 3.9. Average color-naming data obtained for three normal trichromats plotted for wavelengths presented at equal luminance. (after Werner & Wooten, 1979)

100 on the left and the percentage blue or yellow is plotted on the right from 100 to 0. The data could be plotted in this way because, when describing a uniform patch of the visual field, observers do not use the terms red and green simultaneously, that is, they do not call it "reddish green," nor do they use the

terms blue and yellow simultaneously ("bluish yellow"). The arrows in the graphs indicate the wavelengths perceived to be uniquely blue, green, or yellow.

Hering further argued that by studying our color experiences carefully, we could discover other properties of how the brain codes for hue. For example, while we can experience red in combination with either yellow (to produce orange) or blue (to produce violet), we cannot experience red and green at the same time and place. When red and green lights are combined, they cancel each other. The same is true of blue and yellow lights. Hering proposed that this happens because red and green (as blue and yellow) are coded by a single process with two opposing modes of response, excitation and inhibition. A red-green channel can be activated in one direction to signal redness or in the opposite direction to signal greenness, but it cannot simultaneously signal both red and green -- the neural excitation in one cancels the inhibition from the other. Like a seesaw, when one is up the other is down, so red and green cancel each other out. The same holds true for yellow and blue. For this reason, Hering termed red and green, and yellow and blue, ***opponent colors***. Subsequent research on how the brain codes color strongly supports Hering's opponent-colors theory (Zrenner et al, 1990).

As we shall see, the fact that there are a limited number of fundamental hues and that certain color pairs are mutually exclusive can have important practical implications for the appearance of colors in displays. One example can be on a course selector in which the manual radio function is displayed in green and the planned course selector is displayed in magenta. When these two are superimposed, they look white, but the white is coded to mean proposed course modification. In this case, color cancellation on the display may produce confusion.

While the four basic hue terms are sufficient to describe all hues, an account of color appearance must also take into account the achromatic aspects coded by an opponent process that signals black and white. This achromatic channel provides the physiological process for the perception of light and dark colors such as pinks and browns. For example, pink is a bluish red with a substantial white component, and brown is a yellow or yellow red with a substantial black component.

A representation of perceptual color space is shown in Figure 3.10. From our previous discussion, it is apparent that such a representation requires two chromatic dimensions in which red and green are mutually exclusive and yellow and blue are mutually exclusive. In addition, achromatic dimensions must be represented orthogonally to the chromatic dimensions to show the varying degrees of blackness or whiteness in colors.

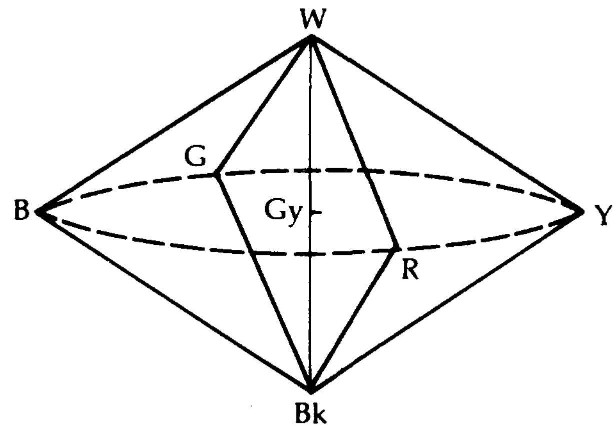


Figure 3.10. Illustration of relations between hue and saturation. (from Hurvich, 1981)

Variations with Intensity

Since the mid-1800s it has been known that as the intensity of a light of fixed spectral composition increases, the hue will change. Specifically, the blue or yellow hue component increases relative to the red or green component. So, for example, as the intensity of a violet light is increased, it will appear more blue than red. This is known as the Bezold-Brücke hue shift. Purdy (1931) quantified this effect and, in addition, reported that three wavelengths, corresponding to the loci of unique hues, were invariant with changes in intensity. There are individual differences in the wavelength of the unique hues.

Figure 3.11 presents data obtained from four observers who were asked to describe the color of a monochromatic light when it was presented at different intensities. The wavelength of the light was 609 nm, which is equivalent to a commonly used red on cathode-ray tube (CRT) displays. Notice that at low light levels, redness is a minor component relative to black and white, but redness and yellowness increase with increasing intensity. Similar results, consistent with a Bezold-Brücke hue shift were obtained for other CRT display colors (Volbrecht et al., 1988).

The data in Figure 3.11 represent a 1° stimulus viewed by the fovea for 1 second. In addition to the loss of hue at low luminances, perception of hue can be further degraded if the stimulus is made smaller and the viewing time is shorter (Kaiser, 1968). When stimuli were presented in a color-naming experiment using small field sizes (less than 15 minutes of arc) and short presentations (50-200 msec), monochromatic or "colored" stimuli were called white 50% of the time (Bouman & Walraven, 1957; Walraven, 1971).

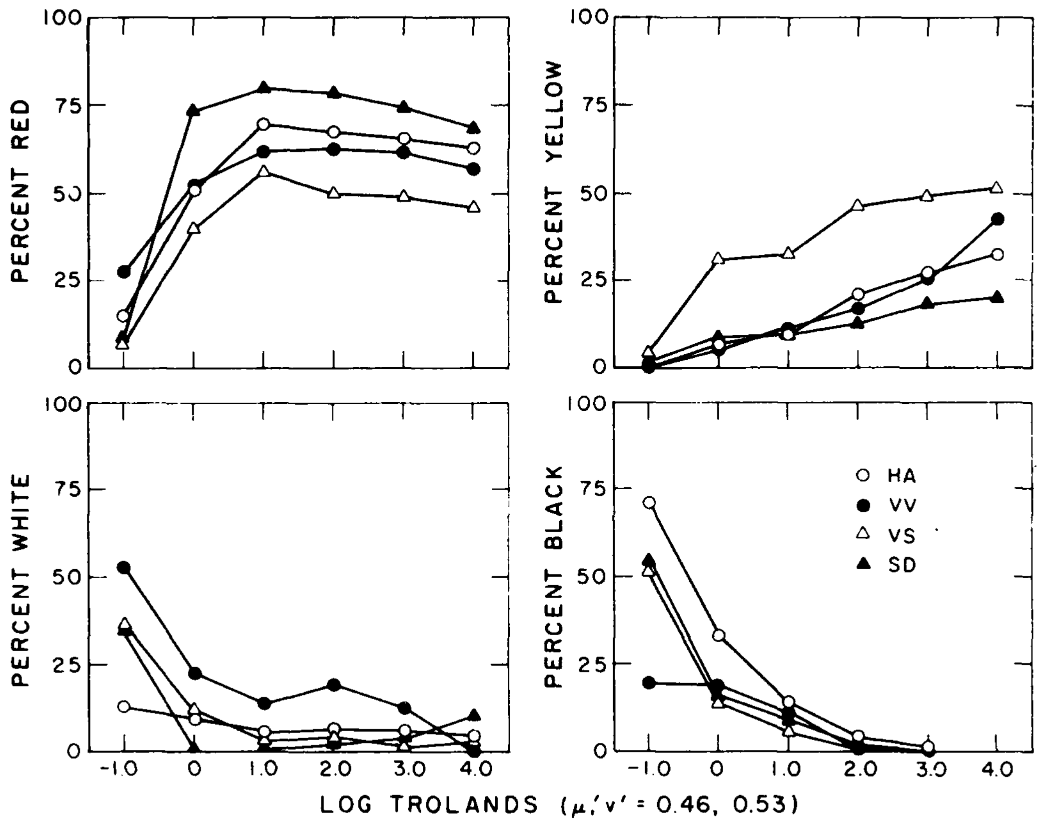


Figure 3.11. Color-naming results plotted as a function of stimulus intensity for four observers. (from Volbrecht et al., 1988)

Variations with Retinal Eccentricity

We have already looked at how the different cone types vary in their distribution with retinal eccentricity. These receptors, of course, provide the input to the neural processes that code the fundamental colors. Thus, it follows that there ought to be some variation in color perception with retinal eccentricity, or with location in the visual field. Sensitivity to color is greatest in the fovea and decreases toward the periphery.

Visual field measurements using stimuli of different color are shown in Figure 3.12. These results are from the right eye of a normal trichromat. The center of the diagram corresponds to the point in the visual field that falls on the fovea and the concentric circles represent positions that move away from the center of the visual field in steps of 10°. The outer, irregularly shaped contour shows the limit of the visual field. Nothing outside this area can be seen with a stationary,

right eye. Inside the visual field are other irregularly shaped contours that define regions in which particular hues can be experienced. Within the central 10° the observer is responsive to all the basic colors: red, green, yellow, blue, black, and white. As we move out from the center, sensitivity to red and green diminishes. Objects that were previously described as reddish yellow and bluish green are now simply seen as yellow or blue. With further eccentricity, the yellow and blue zones diminish and color responses are limited to black and white. Thus, the accuracy with which we can identify colors in a display depends on whether we are looking at them directly or viewing them peripherally.

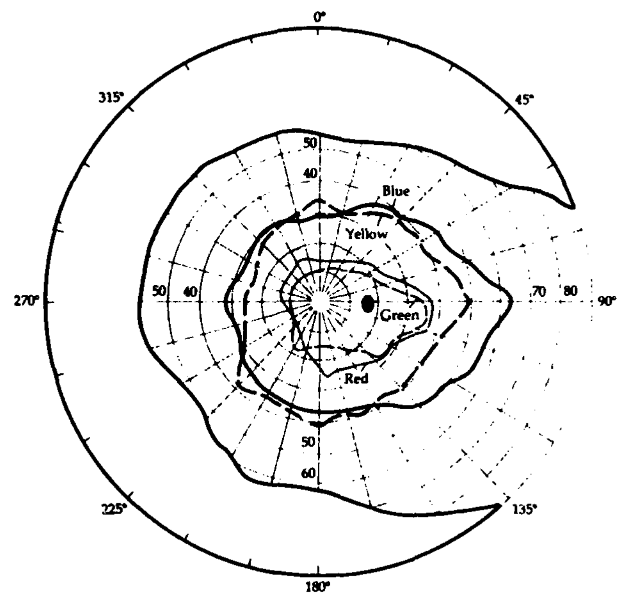


Figure 3.12. Zones in the visual field of the right eye in which various colors can be seen. (from Hurvich, 1981)

There are three points that should be noted about these color zones in the visual field. First, it is evident that the same visual stimulus can be perceived differently depending on the area of visual field that is stimulated. For example, at the fovea, a stimulus might appear orange or reddish yellow, at about 40° away from the fovea it might be yellow, and at 70° it may appear gray. Second, the figure again illustrates that red and green are linked, as are yellow and blue. The linkage is through an opponent code as discussed earlier. Third, these zones were measured under one condition and with other conditions such as larger fields they will change somewhat.

Wavelength Discrimination and Identification

Discriminating color requires an observer to compare two lights and to decide whether they are the same or different. Identification involves an absolute judgment about a color name or category that must be made regardless of whether other colors are present.

Range of Discrimination

To measure wavelength discrimination, the experimenter typically uses a split field such as that shown by the inset of Figure 3.13. One half-field is illuminated by a standard wavelength and the other half-field by a variable wavelength. If the two half-fields are seen as different, the experimenter increases or decreases the intensity of the variable wavelength to determine whether it is discriminable at all intensities. If there is any intensity at which the fields are indiscriminable, it is said that the observer does not discriminate the wavelength pairs. Thus, when we say that two wavelengths can be discriminated, it is implied that this discrimination is made independent of intensity. The object of such an experiment is to find the minimum wavelength difference, or $\Delta\lambda$, that can be discriminated.

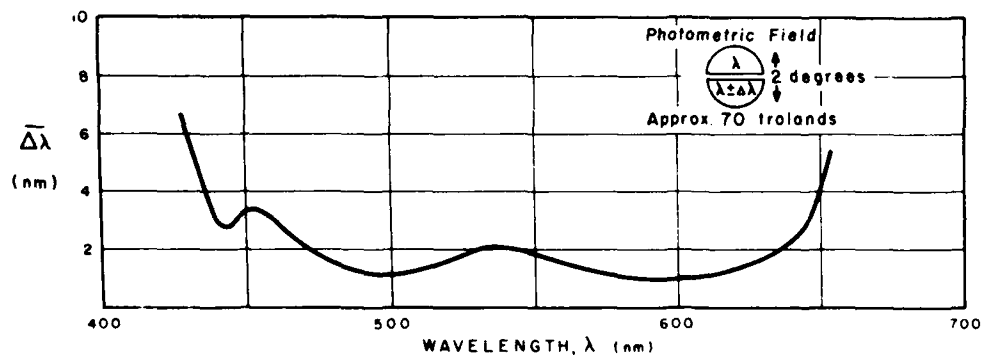


Figure 3.13. Wavelength difference required for discrimination independent of intensity plotted as a function of wavelength. (after Wright & Pitt, 1934)

An average wavelength discrimination function is shown in Figure 3.13. It is plotted as a function of wavelength. There are two minima in the function. At about 500 nm and at about 590 nm some observers can discriminate a wavelength difference of only about 1 nm, regardless of the intensities of the wavelengths. Wavelength discrimination, as with other aspects of color vision, depends on field size. Smaller field sizes are associated with poorer wavelength discrimination (Bedford & Wyszecki, 1958). This means that, all other things being equal, it will be easier to notice a color difference between two relatively large display symbols than two smaller symbols.

The data in Figure 3.13 pertain only to the discriminability of monochromatic lights. To determine the number of discriminable colors requires some account

of nonspectral lights. Based on the number of discriminable hues, number of discriminable steps along the achromatic continuum, and the number of discriminable saturation steps, there are an estimated 7,295,000 color combinations that can be discriminated by the normal human eye (Nickerson & Newhall, 1943).

Range of Identification

According to Chapanis (1965), a set of colors that must be identified on an absolute basis must fulfill several criteria. First, every member of the set must seldom be confused with any other member. Second, every color in the set must be associated with a common color name. Third, use of the color codes should not require specialized training, but should be naturally understood by individuals with normal color vision. To this end, Chapanis asked 40 observers to name 1,359 different color samples (from the Munsell system described on page 69). He then analyzed the data to determine which colors names were used most consistently across observers. Chapanis found that in addition to the achromatic colors (black, white, and gray) which were applied consistently, subjects were most consistent in their use of the terms red, green, yellow, blue, and orange.

Recommendations about the optimum number of colors that ought to be available for visual displays range from about three or four (Murch & Huber, 1982) to ten (Teichner, 1979), the number that can be absolutely identified without extensive training (Ericsson & Faivre, 1988). Use of more than about six or seven colors will lead to errors in identification.

Implications for Color Displays

One often hears of displays that are capable of presenting a large number of colors. In some applications, such as map displays, it may be useful to access a large color palette. However, if colors must be identified, not just discriminated, a large color palette may be of little value. For colors to be identified reliably, they must be distinct under a wide range of viewing conditions. The maximum number that fulfills this requirement is probably not greater than six. Of course, in applications that do not require absolute identification (e.g., cartography), the number of discriminable colors that can be used will increase. The number of colors might also be increased when they are only used to reduce clutter and need not be specifically identified.

In addition to all these considerations, one should heed the conventions for various color choices. For this reason, FAA guidelines (RD-81/38,II, page 50) stress that red should be used for warning indicators and amber for caution

signals. A third color, of unspecified hue, is recommended to indicate advisory level alerts (RD-91/38,II, page 60).

Contrast Effects

The appearance of a color can be altered by another color next to it or another color seen just before or after it. As we scan a scene, we view colors with an eye that has been tuned from moment-to-moment through exposure to preceding and surrounding colors. These contrast effects are dependent on the intensity, duration, and sizes of the stimuli. Here we will illustrate and describe contrast effects, but for detailed summaries of the literature see Graham and Brown (1965) or Jameson and Hurvich (1972).

Successive Contrast

Figure 3.14 illustrates a temporal color-contrast effect. Fixate on one of the dots on the right for a while and then shift your gaze to one of the dots on the white surface to the left. You will see an ***afterimage*** of colors complementary to, that is, opposite, those in the picture. This contrast effect produced over time makes sense if we assume that an opponent-color channel is first driven in one direction by color stimulation and then experiences a rebound effect (of neural activity) in the opposite direction when the stimulus is removed. Thus, we see the opposing color though no external stimulus exists. Wooten (1984) has provided a detailed description of changes in color appearance resulting from successive color contrast.

Simultaneous Contrast

Figure 3.15 illustrates a spatial, color-contrast effect. The thin bars in the two patterns are identical, but they look different when surrounded by different colors. This is called ***simultaneous color contrast*** because it occurs instantaneously. The color induced into the focal area is opposite to that of the surround. This is attributable to opponent processes that operate over space; the neural activity in one region of the retina produces the opponent response in adjacent regions. While the effect noticed here is primarily from the surround altering the appearance of the bars, the opposite also occurs.

Through simultaneous contrast we can experience many colors that are not seen when viewing spectral lights. For example, the color brown is experienced only under conditions of color contrast. If a yellow spot of light is surrounded by a dim white ring of light it will look yellow. As the luminance of the surround is increased (without changing the luminance of the center), there will be

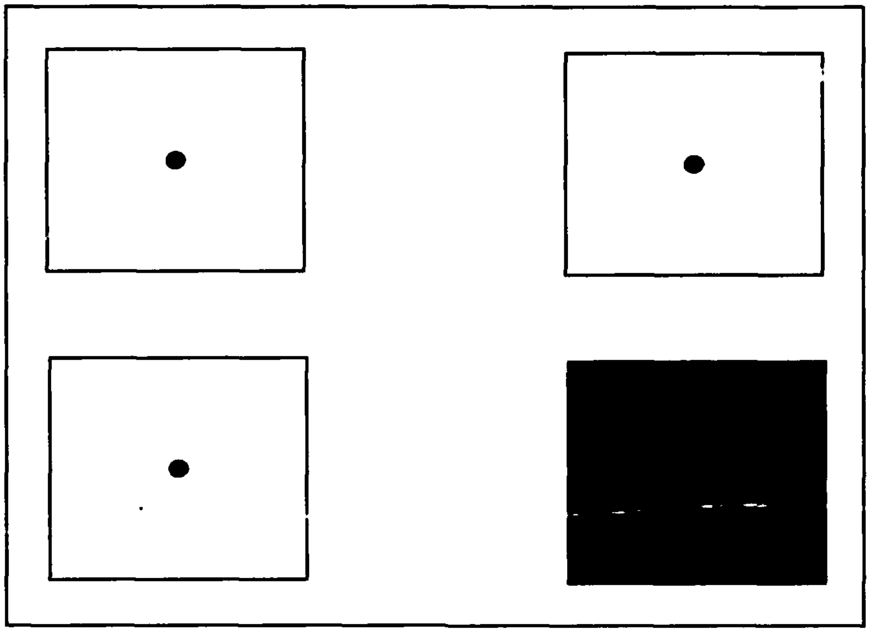


Figure 3.14. A demonstration of successive color contrast (from Hurvich, 1981)

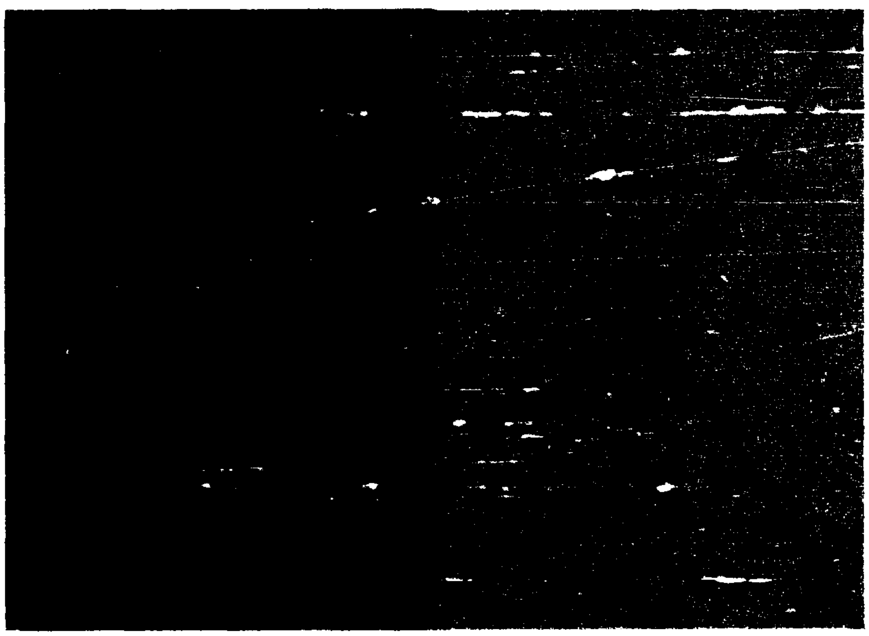


Figure 3.15. A demonstration of simultaneous color contrast (from Albers, 1975)

corresponding changes in the central color. First it will look beige or tan, then light brown, followed by dark brown (Fuld et al., 1983). If the ring is still further increased in luminance, the central spot will look black. The color black is different from the other fundamental colors because it arises only from the indirect influence of light. That is, like brown, the color black is a contrast color and is only perceived under conditions of contrast. Any wavelength can be used in the center or surround and if the luminance ratio is sufficiently high, the center will appear black (Werner et al., 1984).

Assimilation

Sometimes a pattern and background of different colors will not oppose each other as in simultaneous contrast, but will seem to blend together. This is known as ***assimilation*** or the Bezold Spreading Effect and is illustrated by Figure 3.16, (reprinted from Evans, R.M. ***An Introduction to Color.*** Plate XI, p. 192 • John Wiley & Sons, Inc., New York, NY). Here we see that the saturation of the

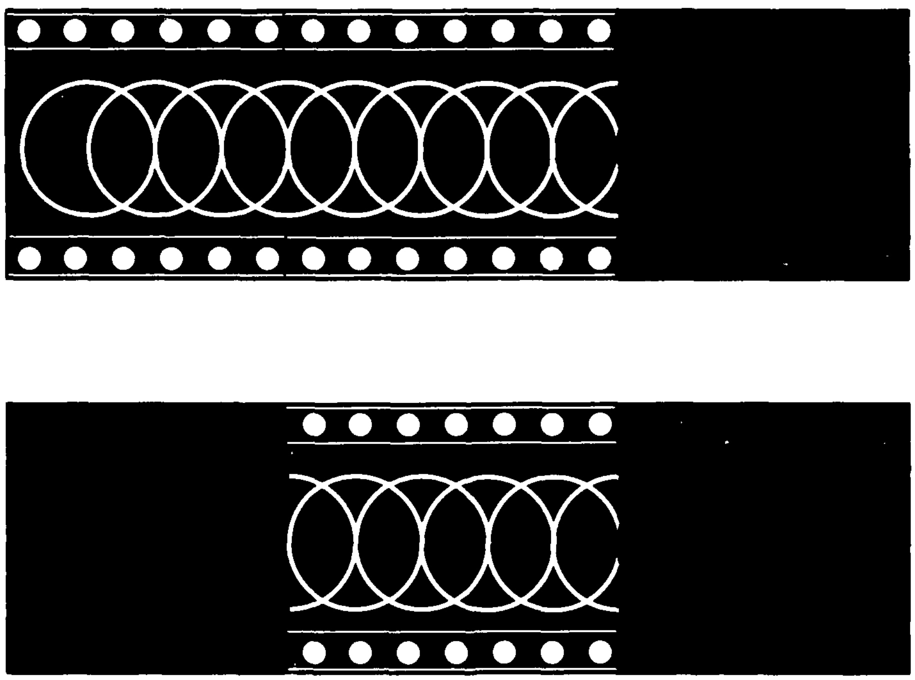


Figure 3.16. A demonstration of assimilation, the Bezold spreading effect (from Evans, 1948)

red background of the top left and center looks different depending on whether it is interlaced with white or black patterns, even though the background is physically the same in the two sections. The lower illustration shows the effect

of assimilation with a blue background. Assimilation is not well understood, but it is known that it cannot be explained by light scatter from one region of the image to another. The phenomenon arises from the way in which colors are processed by the brain.

Adaptation

We have already seen from the dark adaptation curve that the visual system changes its sensitivity according to the surrounding level of illumination. We have also seen that visual acuity increases with increased light level. Here we shall briefly discuss some of the changes in color perception that occur with changes in ambient light.

Chromatic Adaptation

The appearance of a color can be altered by preceding or surrounding colors that are only momentarily in the field of view. Even larger effects can be observed when an individual is fully adapted to a chromatic background. This is demonstrated by an experiment of Werner and Walraven (1982) in which the subject was instructed to adjust the ratio of two lights so that the mixture would appear pure white. The subject then viewed an 8° chromatic adapting background for seven minutes and again adjusted the ratio of the two lights so that it looked white. The results are shown in Figure 3.17 using the CIE color diagram that will be explained below. For now, consider that the color diagram represents all mixtures of colors. The central x designates the mixture that appeared white in the neutral state (dark background) and the lines radiating outward connect the neutral white point with the chromaticity of the adapting background (on the perimeter of the diagram). The individual data points show the light mixture that appeared white after chromatic adaptation. You can see that the light mixture that appears white is dramatically altered by chromatic adaptation.

In part of the experiment, the intensity of the chromatic background was kept constant, but the intensity of the test spot was varied. Contrast refers to the ratio of the increment to the background. The results show that lower contrasts are associated with larger shifts in the white point. Indeed, nearly any light mixture can appear white under the appropriate conditions of adaptation and contrast.

In natural settings one does not ordinarily adjust the chromaticity of a stimulus to maintain a constant color, although devices to implement such a scheme on aircraft displays have been proposed (Kuo & Kalmanash, 1984). What ordinarily happens is that adaptation alters the color of a stimulus in a direction opposite

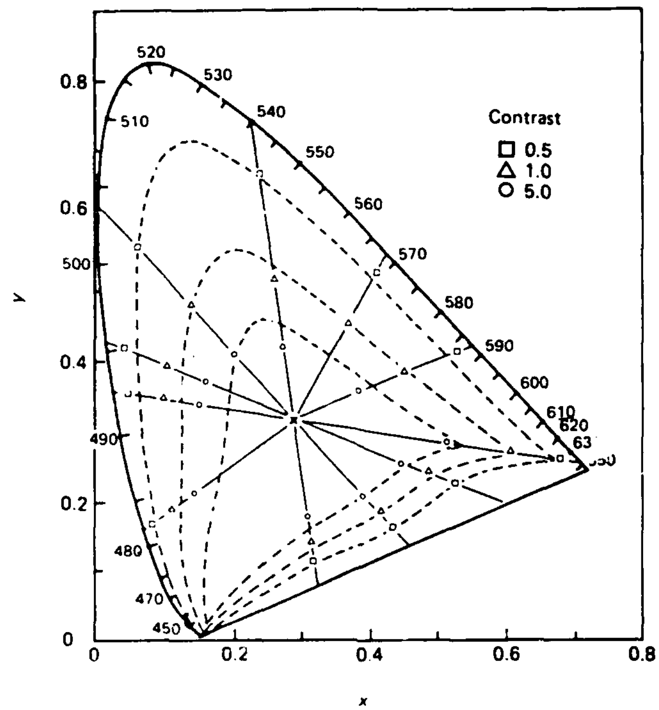


Figure 3.17. Chromaticity diagram showing stimuli that appear white under dark-adapted condition (central x) and following adaptation to chromatic backgrounds (filled circles). (after Werner & Walraven, 1982)

to that of the adapting field color. For example, white letters may be tinged with yellow when viewed on a blue background or tinged with green when the observer has adapted to a red background. These effects of chromatic adaptation can be altered to work in favor of color identification or detection. For example, detection of a yellow stimulus may be enhanced by presenting it on a blue background.

Variation Under Normal Conditions

The effects of ambient light in altering the state of adaptation are not fundamentally different from those already shown in Figure 3.17. However, since most ambient lights contain a broad distribution of wavelengths, the receptors are not adapted as selectively as in laboratory experiments.
One important consideration in evaluating changes in ambient illumination under natural conditions is that in addition to altering the perceptual state of an observer, there often can be substantial changes in the display itself. CRT screens typically reflect a high percentage of incident light. The light emitted

from the display is therefore seen against this background of ambient light. Figure 3.18 shows how sunlight alters the spectral composition of the colors available on a display. As sunlight is added to the display, the gamut of chromaticities shrinks, as illustrated by the progressively smaller triangles (Viveash & Laycock, 1983). To an observer this would be experienced as a desaturation or "wash out" of the display colors as well as a shift in hue that accompanies changes in saturation, called the Abney effect (see Kurtenbach, Sternheim & Spillmann, 1984). Some colors that were previously discriminable may no longer be so. Finally, not illustrated by the figure is the substantial reduction in luminance contrast with increasing ambient illumination. Some visual displays on aircraft are automatically adjusted in their luminance by sensors that respond to the ambient illumination (e.g., all CRTs on Boeing 757 and 767). This is an important innovation, and indeed consistent with FAA recommendations (RD-81/38,II, page 47) that alerting signals be automatically adjusted according to the ambient illumination level. However, manual override control is also recommended (RD-81/38,II, page 73) to compensate for individual differences in sensitivity, adaptation, and other factors such as use of sunglasses.

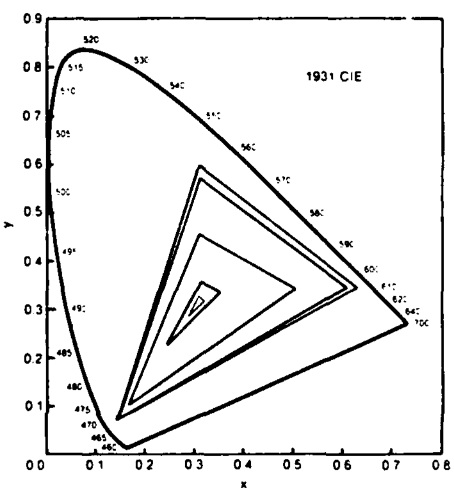


Figure 3.18. Chromaticity diagram showing how the color gamut of a display decreases with increasing sunlight. (after Viveash & Laycock, 1983)

Color Specification

There are many situations in which it is useful to have an objective method for specifying color. Since color perception of a fixed spectral distribution depends upon many conditions, a system of color specification could be based on appearance or on some physical or psychophysical description of the stimulus. Each system of color specification has advantages and disadvantages.

CIE System

We have seen that a normal trichromat can match any wavelength (or any mixture of wavelengths) by some combination of three other wavelengths or

primaries. The choice of wavelengths for the primaries is somewhat arbitrary, but different sets of primaries will necessarily involve different intensity ratios.

Since any color can be matched by some mixture of three primaries, any color can be represented in terms of the proportional contribution of each primary to the mixture. For example, a light that is matched with 10 units of wavelength 450 nm, 5 units of 550 nm, and 20 units of 670 nm has a ratio of the three primaries of 2:1:4. While our ratio of primaries would provide an exact match to the light of interest, other primaries could also be used to provide an exact match. To be useful in a wide variety of applications, it would be helpful if specifications of a color could all be made in terms of the same set of primaries. Thus, in 1931 the CIE developed a set of imaginary primaries to represent the color-matching functions for a standard observer. Since these primaries are not real, they are given the arbitrary labels X, Y, and Z. Figure 3.19 shows the relative amount of these theoretical primaries needed to match any wavelength of unit energy. The values plotted here are designated $\overline{X}$, $\overline{Y}$, and $\overline{Z}$, and are known as the spectral tristimulus values. Among the nuances of this system, the $\overline{Y}$ tristimulus value is identical to the V_λ function (photopic sensitivity of the standard observer). Thus, when the $\overline{Y}$ tristimulus value is integrated with the energy distribution (by multiplying the energy by $\overline{Y}$ at each wavelength and summing), we have the total value of the Y primary which is equal to the luminance. It should also be mentioned that the CIE actually developed two sets of tristimulus values, one for 2° stimuli and one for 10° stimuli.

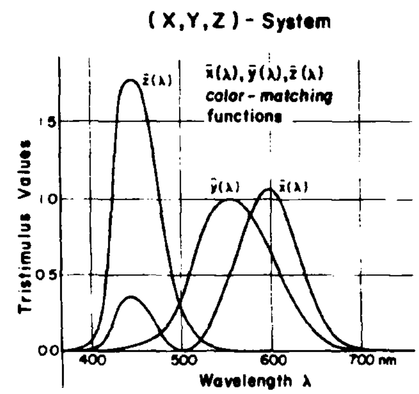


Figure 3.19. CIE tristimulus values for a 2° standard observer plotted as a function of wavelength. (from Wyszecki & Stiles, 1982)

To specify the chromaticity of a particular color in the CIE system, the energy at each wavelength is multiplied by the $\overline{X}$ tristimulus value at each wavelength and the products are summed across wavelengths to yield the tristimulus value (not to be confused with the spectral tristimulus values) designated as X. Similarly, the energy across wavelengths is convolved with the $\overline{Y}$ and $\overline{Z}$ tristimulus values to yield Y and Z. The X, Y, Z values can be quite useful in specifying a color. For example, given the values for a color of interest, we can be certain that it

can be matched with respect to the standard observer by an individual who creates these same X, Y, Z values using any other wavelength combination.

We now have the ingredients for representing a color in question in an x,y chromaticity diagram that represents all conceivable colors. The chromaticity coordinates are defined as: $x = X/(X + Y + Z)$; $y = Y/(X + Y + Z)$; $z = Z/(X + Y + Z)$. Notice that x, y, and z are proportions that sum to 1.0. Thus, it is only necessary to plot x and y since $z = 1 - (x + y)$. The resulting chromaticity diagram is shown in Figure 3.20. Notice that monochromatic lights

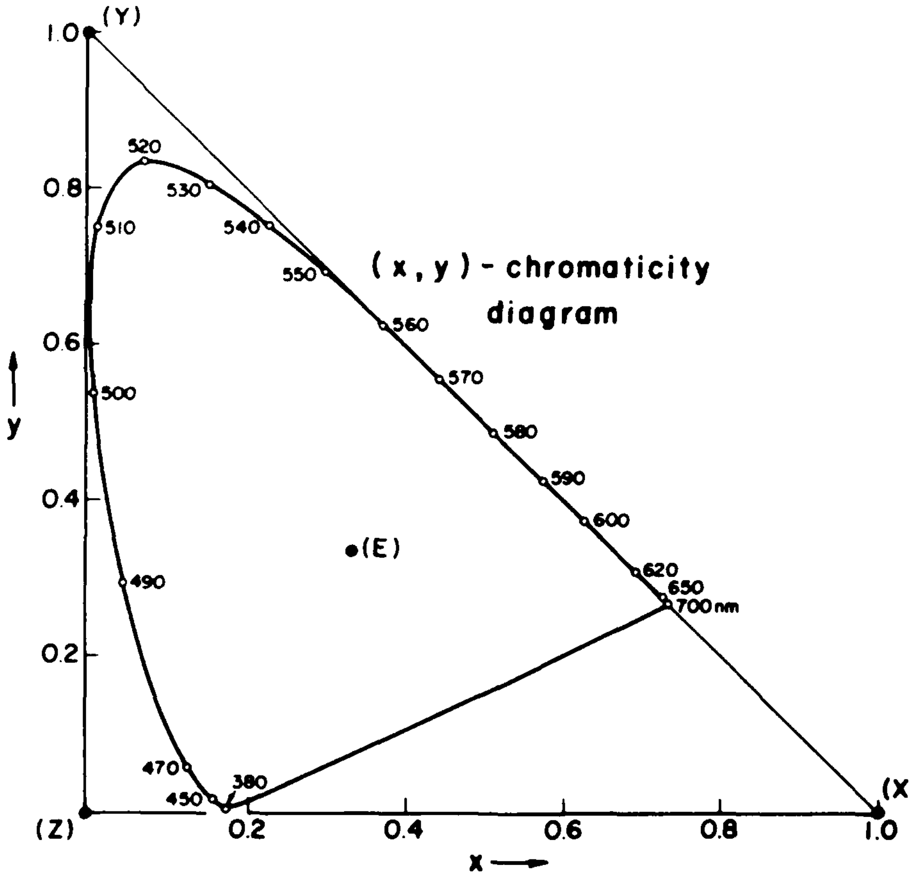


Figure 3.20. CIE color diagram. (from Wyszecki & Stiles, 1982)

all plot around the perimeter of the diagram, a region known as the spectrum locus. The area inside the diagram represents all physically realizable mixtures of color. Given the chromaticity coordinates of a color, a perfect match can be made by various mixtures determined using the chromaticity diagram. If we also wanted the match to include information about luminance, we would have to specify Y as well as the x,y coordinates.

A useful property of the CIE chromaticity diagram stems from the fact that a mixture of two lights always plots on a straight line that connects the points representing the lights within the diagram. The position along the line that represents the mixture depends on the energy ratio of the two lights. Thus, if we plot the points representing the chromaticity coordinates of three phosphors on a color display, we can connect the points to create a triangle representing the color gamut of the display. This triangle would represent all chromaticities that can be generated by the display.

The CIE chromaticity diagram is useful for specifying color in many applications, but it does have some drawbacks. Perhaps the most important problem is that equal distances between sets of points in the diagram are not necessarily equal distances in perceptual space. To rectify this problem the CIE developed a new chromaticity diagram, shown in Figure 3.21, in an attempt to provide more uniform color spacing. The coordinates of this diagram are called u',v' and can be obtained by a simple transformation from the x,y coordinates (for further details see Wyszecki & Stiles, 1982). The smaller triangle in Figure 3.21 shows the gamut of many typically used displays while the larger triangle shows the maximum envelope of currently used displays.

Munsell System

The CIE system is useful for specifying the chromaticity of a visual stimulus, but no information about color appearance is preserved. The appearance of lights of a fixed chromaticity will depend on many variables, as was illustrated in Figure 3.17. Several systems for specifying color that are easier to use and more closely related to perception than the CIE system are available, perhaps the best

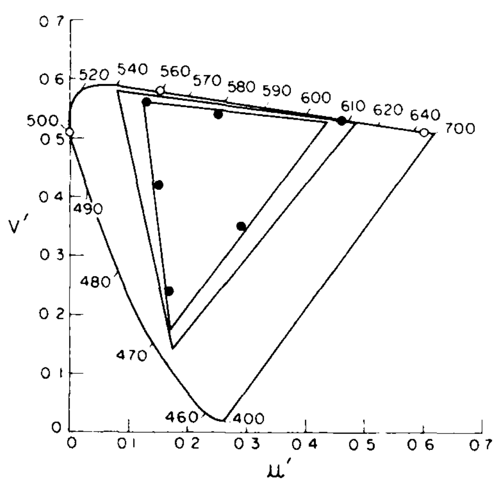


Figure 3.21. **CIE, u', v' chromaticity diagram based on a 2° standard observer. (from Volbrecht et al., 1988)**

known system being the one developed by Munsell in 1905. In its current form, the Munsell System consists of a series of colored paint chips arranged in an orderly array, as illustrated by Figure 3.22. Each entry is characterized by three numbers that specify hue, blackness and whiteness from 0 to 10 (called lightness), and ratio of chromatic and achromatic content (called chroma). Hue is represented by a circular arrangement in 40 steps that are intended to be equal in perceptual space. Lightness varies from bottom to top in nine equally spaced steps from black to white. Chroma, or saturation, represents the hue and lightness ratios in 16 steps that vary from the center outward. To use this system, one merely finds the chip that most closely matches an item of interest. Each chip is specified by three parameters: hue, lightness, and chroma. Since the steps between chips are nearly equal, the Munsell system can be useful in the selection of colors that are equal distances in perceptual space.

While the Munsell system is easy to use and the arrangement corresponds more closely to color appearance than the CIE system, it still has many limitations. The influences of surrounding colors and state of adaptation which are important for color appearance are not taken into account by the Munsell designations. Thus, the appearance and discriminability of colors expected from a Munsell designation may not be obtained when the conditions of viewing are altered.

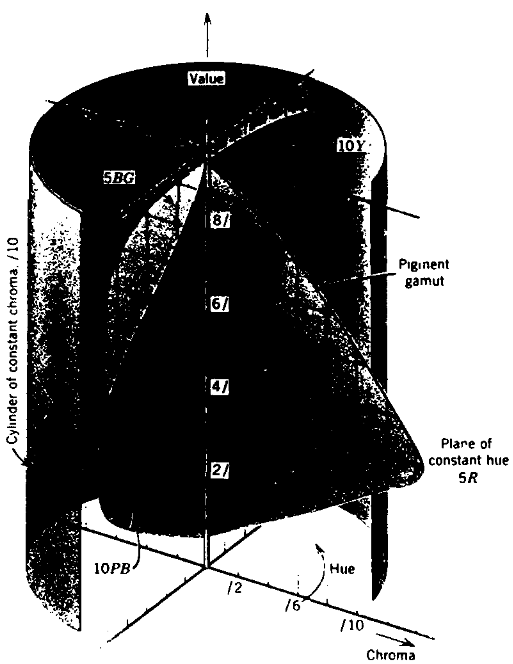


Figure 3.22. Schematic of the Munsell color solid. (from Wyszecki & Stiles, 1982)

Implications for Displays

Color can significantly enhance search and identification of information on visual displays. It is more effective than shape or size in helping to locate information quickly (Christ, 1975). The attention-getting nature of color facilitates search while at the same time providing a good basis for grouping or organizing information on a display which may help display operators segregate multiple types of information and reduce clutter. For example, an experiment by Carter (1979) showed that when the number of display items was increased from 30 to 60, search time increased

by 108% when only one color was used, but increased by only 17% for redundant color-coded displays.

There are severe constraints on the effective usage of color information (see also Walraven, 1985). The attention-getting value of a color is dependent on its being used sparingly. Only a limited number of colors should be used in order to avoid overtaxing the ability of an observer to classify colors. If each color is to have meaning, only about six or seven can be utilized effectively.

In addition, we have seen that perception of a fixed stimulus will be changed as a function of many variables including the intensity, surrounding conditions, temporal parameters, and state of adaptation of an observer. If color is a redundant code, these problems, as well as loss of color due to aging of the display, will have substantially less impact on operator performance.

The choice of colors can be facilitated by considering the physiological principles by which hues are coded -- red opposes green and blue opposes yellow. These colors are also separated well in CIE chromaticity diagrams. Colors that are barely discriminable at low ambient conditions may not be at all discriminable at high ambient conditions because of a physical change in the color gamut.

The use of blue stimuli can be problematic for displaying characters requiring good resolution. The blue phosphors on many displays only produce relatively low luminances, but the main difficulty is a physiological problem in processing short wavelengths. One problem already mentioned that might result from using small blue stimuli is related to small-field tritanopia. Because the short-wave cones are distributed more sparsely across the retina, they contribute very little to detail vision. Short-wave cone signals are not used in defining borders or contours (Boynton, 1978). In addition, focusing of short-wavelength stimuli is not as easily achieved as for middle- and long-wave stimuli, making blue a color to avoid in displaying thin lines and small symbols. A major advantage of blue and yellow is that our sensitivity to these colors extends further out in the visual field than our sensitivity to red and green. Blue hues also provide good contrast with yellow. Thus, while blue may be a good color to avoid when legibility is a consideration, it may be a good color to use for certain backgrounds on displays.

Chapter 4

Form and Depth

by John S. Werner, Ph.D., University of Colorado at Boulder

We have already discussed how two objects of different sizes, placed at different distances from us can cast images of identical size and shape on our retina. Despite this, we can still tell that one is small and close and the other is large and far away. How do we do this? Either we have additional information about physical distance or we know something about the physical size.
We encounter another aspect of the same perceptual problem when we consider the fact that as an object changes position with respect to us, because either it is moving or we are moving, the retinal image formed by the object continuously changes shape and size. These changes depend on both the object's distance and our angle of view. For example, an object moving away "grows" smaller. Or the image of a square on our retina may become in turn a rectangle or a trapezoid depending on our angle of view. The amazing fact in the face of such retinal contortions is that our perceptions of the object's shape and size remain relatively constant; we still see a square. These perceptual

constancies, termed ***shape*** and ***size constancy***, require information about distance, not only the distance of objects in relation to each other, but also the distances between points on the same object, say the corners of a square. Somehow we process the information we take in through our retinas at higher levels in the nervous system in terms of information we hold about size, shape, distance -- in other words, our concepts about the physical world. It is important to realize that we are usually unaware of this process when perceiving size and distance; we do it automatically. In this section we will discuss some of the ways in which form and depth information are processed.

Edges and Borders

The eye movement records shown before in Figure 2.17 suggested that borders and edges of a stimulus were often the target of visual fixation. Our ability to separate figure and ground in a complex scene requires differences in light level. It is not the overall light level that typically defines an object's edge or border, it is a difference in light levels -- the contrast. There are several ways to define contrast, but in this section we will define it as:

$$(L_{max} - L_{min}) \,/\, (L_{max} + L_{min})$$

where L_{max} is the maximum luminance in the pattern and L_{min} is the minimum luminance in the pattern. With this definition, contrast can vary between 0 and 1.0.

The importance of contrast in defining the brightness or lightness of an object or area was already illustrated by simultaneous brightness contrast. The brightness of a point of light within a pattern is partly determined by its own characteristics but also by the brightness of points surrounding it. Many of the processes responsible for simultaneous contrast originate within the retina. The information that retinal cells send to the brain has little to do with overall light level, rather they are coding small differences in light level from one region to the next.

A striking consequence of the way in which the visual system extracts brightness information is illustrated by Figure 4.1. The top panel represents a black-and-white disk that can be mounted to a motor and spun rapidly. The black region reflects about 5% of the light falling on it and the white region reflects about 85%. Now imagine that the disk is spun rapidly so that you cannot discern the separate black and white regions. This is shown by the middle panel. If we measure the light reflected from the disk by passing a small probe from left to right, the intensity of light would vary with the ratio of black-to-white areas. The bottom panel shows a luminance profile of this

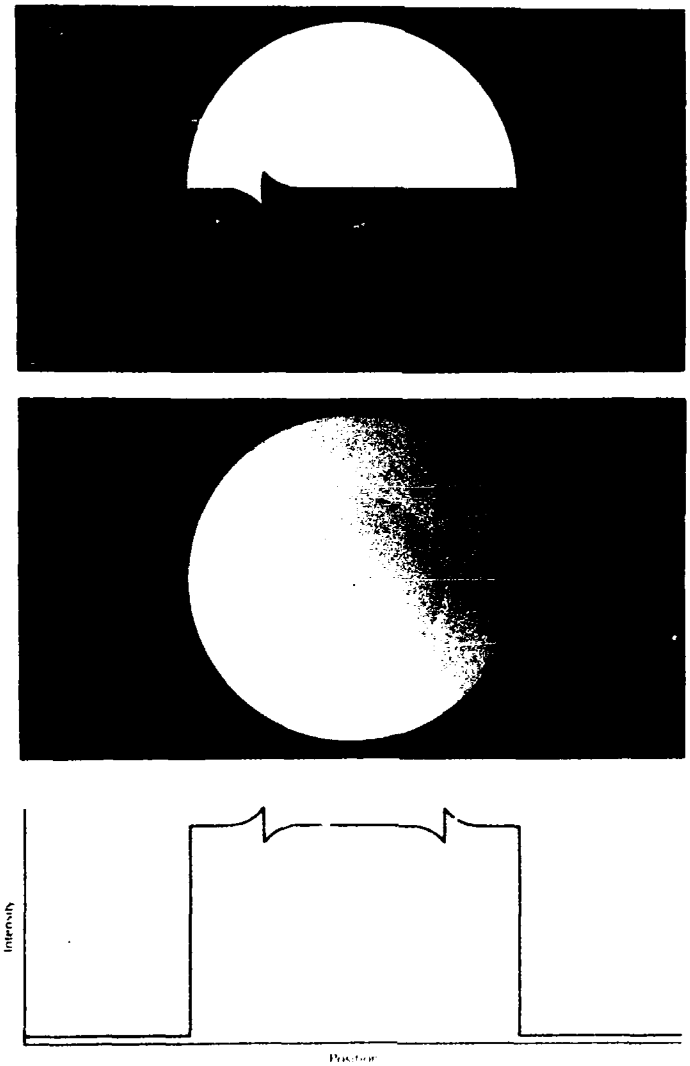


Figure 4.1. Demonstration of the Craik-Cornsweet-O'Brien illusion. (from Cornsweet, 1970)

stimulus; that is, a graph of the light intensity or luminance plotted as a function of spatial position. Notice that the inside and outside of the pattern are separated by a change in light level, or border, but beyond this change they have the same black-to-white ratio and hence the same luminance is reflected to the eye. If brightness depended on light intensity alone, these two regions should be perceived as identical. This, however, is not what we perceive; the inside region is perceived as darker than the outside region. This effect is known as the Craik-Cornsweet-O'Brien illusion. It shows that the brightness of a region of light is dependent on the contrast at the border.

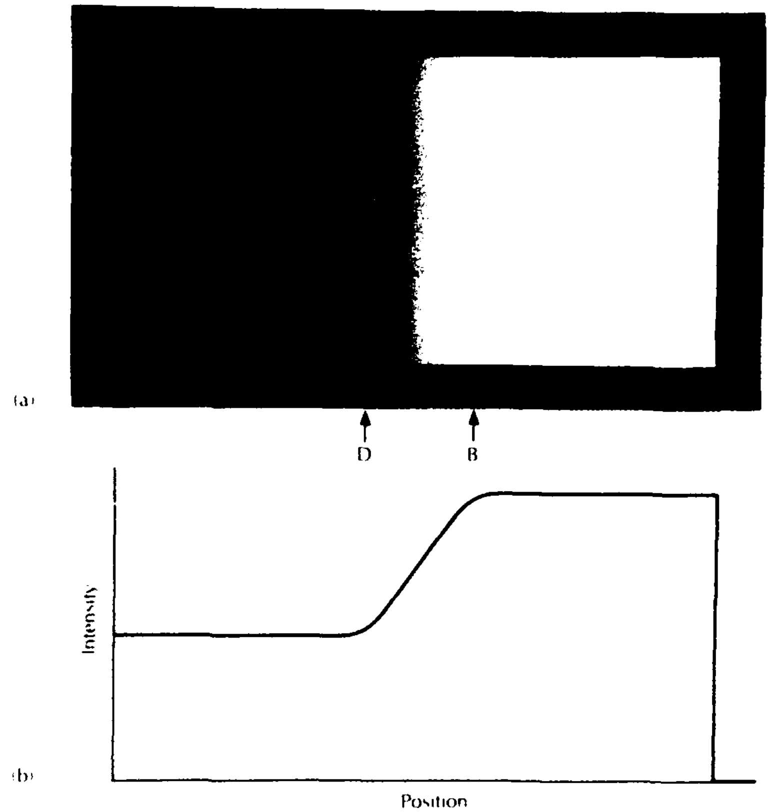


Figure 4.2. Illustration of Mach bands. (from Cornsweet, 1970)

The bottom panel of Figure 4.2 shows a luminance profile in which there is an increase in intensity from left to right. The photograph above shows a stimulus that changes according to this luminance distribution, but notice that our perception does not follow it exactly. Rather, at the border one perceives small bands of exaggerated darkness and brightness, labelled D and B in the photograph. These are called ***Mach bands*** in honor of Ernst Mach (1865) who first described them. The pattern we perceive exaggerates the abrupt light-dark transitions.

There are many other phenomena in which the brightness or darkness of a region depends on border contrast or on changes in contrast over time (see Fiorentini et al., 1990). These phenomena reveal the visual system's attempt to extract information at the borders because borders and edges define objects or parts of objects.

Contrast Sensitivity

The forms of objects are defined by contrast. It is, therefore, important to characterize the sensitivity of the visual system to contrast. One approach to this problem is to measure contrast sensitivity using grating stimuli in which the luminance is varied sinusoidally as illustrated by Figure 4.3. If one were to

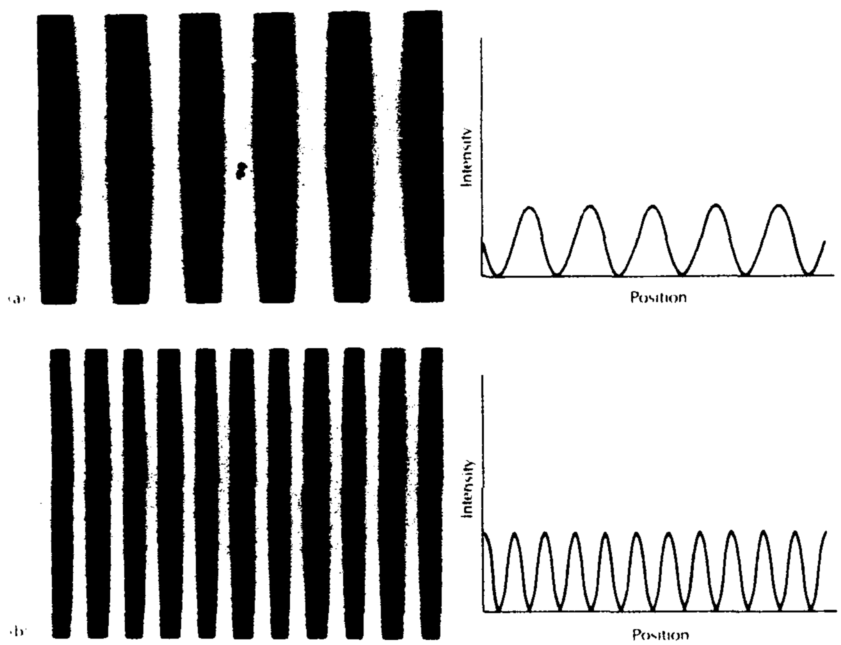


Figure 4.3. **Vertical sine-wave gratings and their luminance distributions. (from Cornsweet, 1970)**

measure the intensity of the stimuli on the left, by passing a light meter across it, the sinusoidal luminance profile on the right would be found. The profile of the stimuli could be characterized by the contrast, which was defined above by the difference between the luminance maximum and minimum, divided by the average luminance. The frequency of oscillation of the sine wave is defined in terms of the number of cycles per degree of visual angle (cpd). For example, the stimulus on the top of Figure 4.4 has a lower spatial frequency than the one on the bottom.

Contrast threshold is measured by determining the minimum contrast required for detection of a grating having a particular spatial frequency (usually generated on a CRT display). Contrast sensitivity is the reciprocal of contrast threshold. Thus, the ***contrast sensitivity function*** represents the sensitivity of an

individual to sine-wave gratings plotted as a function of their spatial frequency. Figure 4.4 shows a typical contrast sensitivity function (Campbell & Robson, 1968). These data were obtained with a set of static sine-wave gratings (like those in Figure 4.3), but contrast sensiti... functions vary as a function of luminance, temporal characteristics of the grating stimuli (e.g., flickering vs. steady), and stimulus motion characteristics (e.g., drifting vs. stationary gratings). The shape of the contrast sensitivity function also varies with the individual observer and the orientation of the grating. For example, many individuals are more sensitive to vertical and horizontal gratings of high spatial frequency than to oblique (45° or 135° from horizontal) gratings (Appelle, 1972).

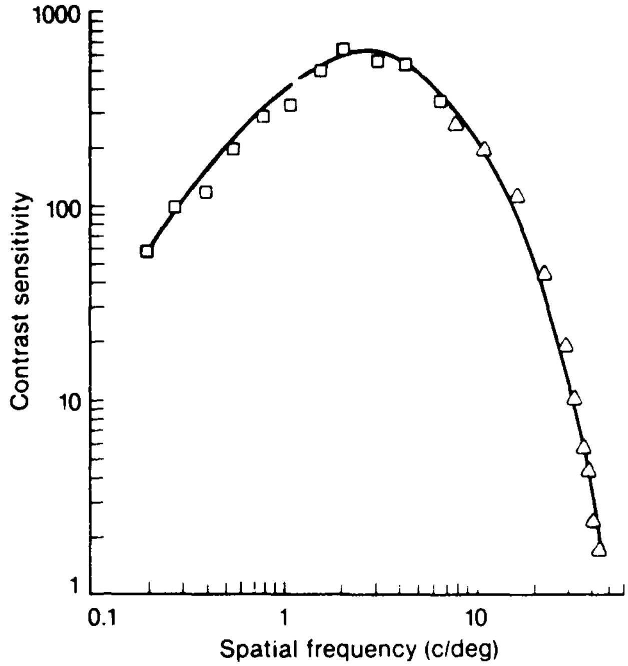


Figure 4.4. Contrast sensitivity as a function of spatial frequency. (from Campbell & Robson, 1968)

It can be deduced from the contrast sensitivity function that we are not equally sensitive to the contrast of objects of all sizes. High spatial frequency sensitivity is related to visual acuity; both are a measure of resolution, or the finest detail that can be seen. When spatial vision is measured by an optometrist or ophthalmologist, only visual acuity is typically measured. While a more complete evaluation of spatial vision would include contrast sensitivity measurements over a range of spatial frequencies, it is the high frequency sensitivity that is most impaired by optical blur (Westheimer, 1964). Thus, high frequency sensitivity is what is improved by spectacle corrections.

One explanation for our contrast sensitivity is that cells in the visual cortex respond selectively to a small band of spatial frequencies. The contrast sensitivity function may thus represent the envelope of sensitivity of these cells. This is analogous to the photopic spectral sensitivity function representing the relative activity of three classes of cones. In the case of contrast sensitivity, the model implies that different cells respond selectively to stimuli of different sizes. A demonstration consistent with this idea is presented in Figure 4.5. Notice that

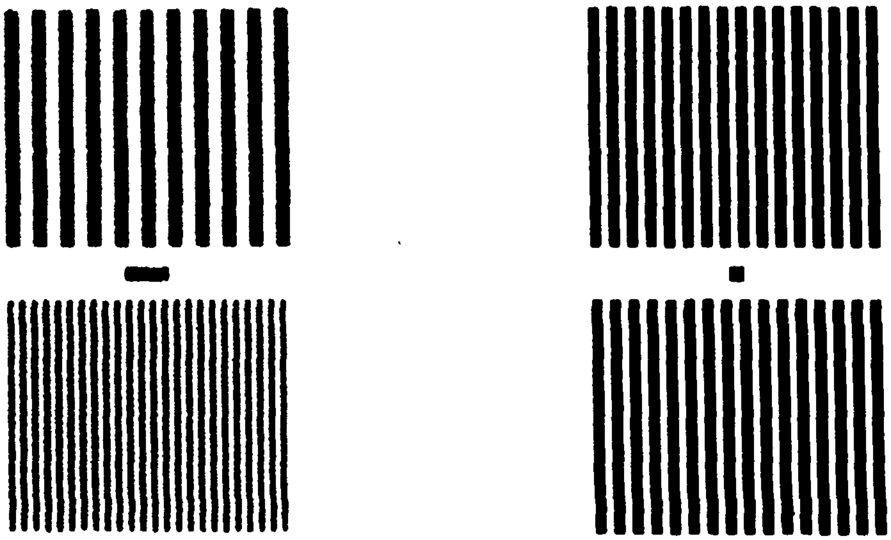


Figure 4.5. Demonstration of size-selective adaptation. (from Blakemore & Sutton, 1969)

the two patterns on the right are of identical spatial frequency. Now, stare at the bar between the two gratings on the left, allowing your eyes to move back and forth along the bar. This scanning prevents the buildup of a traditional afterimage. It is intended to fatigue cells responsive to gratings of a particular size. After about 45 seconds of fixating along the bar on the left, shift your gaze to the small bar on the right. The two patterns on the right will now appear to have different spatial frequencies. According to theory (Blakemore & Sutton, 1969), size-selective cells responsive to gratings on the left were fatigued during fixation. This shifted the balance of activity when looking at the patterns on the right compared to the activity produced by the gratings prior to adaptation.

Variation with Luminance

The effects on contrast sensitivity of changing the space average luminance of the stimulus were systematically investigated by DeValois, Morgan and Snodderly (1979). Their data are shown in Figure 4.6. Contrast sensitivity is plotted as a function of spatial frequency. Different symbols and curves from top to bottom correspond to luminance decreases in steps of 1.0 log unit. This figure shows that overall contrast sensitivity is reduced as luminance decreases, but the reduction in sensitivity is much greater for high than low spatial frequencies. This shifts the peak of the function to lower frequencies with reduced luminance. In general, high spatial frequency sensitivity decreases as a function of the square root of the luminance (Kelly, 1972).

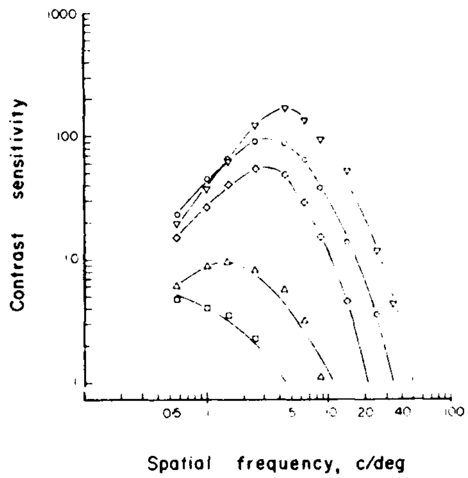


Figure 4.6. Contrast sensitivity is plotted as a function of spatial frequency for young, adult observers. (from DeValois, Morgan & Snodderly, 1974)

Variation with Retinal Eccentricity

As we have seen, the peak of the spatial contrast sensitivity function occurs at about 3-5 cpd at moderately high luminances and declines at lower and higher frequencies around the peak. When the same function is measured at different retinal eccentricities using a display of fixed size, the results depend strongly on distance from the fovea, as shown by the panel on the left of Figure 4.7 (Rovamo, Virsu & Näsänen, 1978). Measurements were obtained with a 1° x 2° vertical grating. Different curves refer to different retinal eccentricities; from the highest curve down these were 0°, 1.5°, 4.0°, 7.5°, 14°, and 30° from the fovea. There are a number of reasons for this dependence on eccentricity, including the variation in receptor distribution with eccentricity and the way in which receptor signals are pooled in the retina and at higher levels in the brain (see Wilson et al., 1990). When larger stimuli were used to compensate for these factors, Rovamo et al. obtained the results shown in the panel on the right side of Figure 4.7. These results are important because they show how stimuli can be scaled in size to be equally visible at all eccentricities.

Variation with Age

Average contrast sensitivity for various spatial frequencies are shown in Figure 4.8 plotted as a function of age. These data represent averages from 91 clinically normal, refracted observers tested by Owsley, Sekuler and Siemsen (1983). Age-related declines in contrast sensitivity, like declines related to decreased retinal illuminance, are most pronounced at high spatial frequencies (see also Higgins, Jaffe, Caruso & deMonasterio, 1988). These findings are consistent with studies which examined the relation between age and static visual acuity (Pitts, 1982), a measure that is primarily dependent on the transmission of high spatial frequencies, and known to decline with advancing age. Because the lens transmits less light and the pupil is smaller in elderly observers, the change in contrast sensitivity may be partly a luminance effect, although changes in neural structures also play a role.

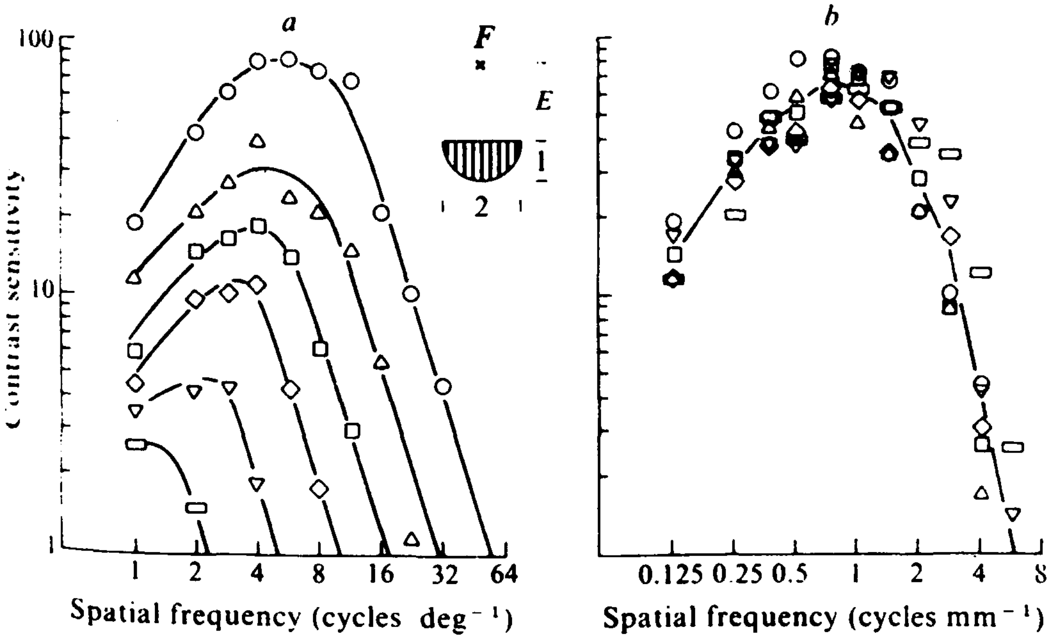


Figure 4.7. Contrast sensitivity measured at different retinal eccentricities is plotted in the graph on the left as a function of spatial frequency. The graph on the right shows contrast sensitivity obtained at the same retinal eccentricites but with a stimulus that was scaled according to "neural" coordinates. (from Rovamo, Virsu, & Näsänen, 1978)

Implications for Displays

The contrast sensitivity function has several areas of application. First, as a predictor of visual performance, the contrast sensitivity function may be more useful than traditional measures of visual acuity. The visual acuity chart varies only the size of the stimuli to evaluate spatial vision while contrast sensitivity testing requires variation in both size and contrast. The importance of this additional information about contrast was demonstrated by Ginsburg et al. (1982). They conducted an experiment with experienced pilots and an aircraft simulator. The simulated visibility was poor and half of the simulated landings had to be aborted due to an obstacle placed on the runway. Performance was measured by how close the pilots flew to the obstacle before aborting the landing. Pilot responses on this task (times required to abort the landing) varied considerably. Individual differences in performance were not well correlated with visual acuity, but were well predicted by individual variation in contrast sensitivity. Thus, contrast sensitivity testing may be more useful than traditional measures of visual performance for predicting responses in complex settings.

A second application of the contrast sensitivity function is for predicting the visibility of complex patterns presented on displays. It may not be feasible to test every unit of symbology directly, but knowing the contrast sensitivity function, it may be possible to make some predictions using a spatial frequency analysis of the stimulus. This approach is based on Fourier's theorem, according to which any complex waveform can be described in terms of a set of sinusoidal waves of known frequency, amplitude and phase (the alignment of the waves when added together). This approach was briefly introduced (page 2) when discussing the processing of complex tones by decomposing them into individual pure tones, and it was used to predict temporal sensitivity to complex waveforms on the basis of sensitivity to sinusoidal waves.

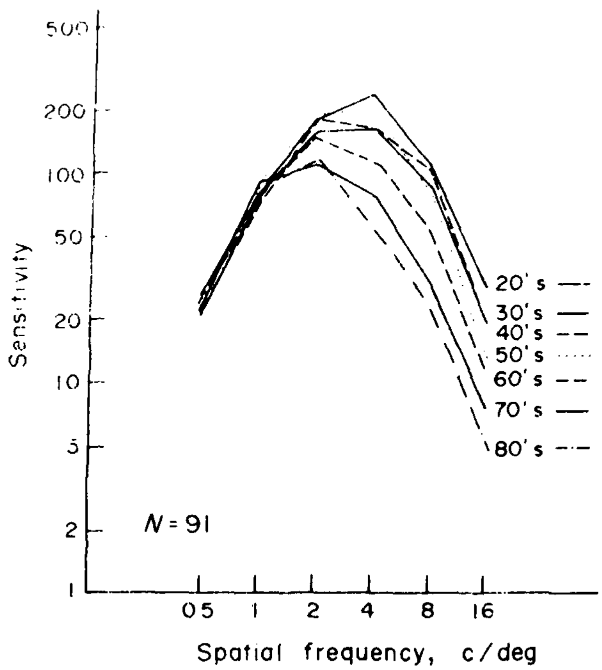


Figure 4.8 Contrast sensitivity as a function of spatial frequency for different age groups. (from Owsley, Sekuler & Siemsen, 1983)

To illustrate how a complex pattern can be described in terms of a set of sine waves, consider the difference between a sine-wave grating and a square-wave grating. Figure 4.9 shows these two types of grating at the same spatial frequency. The square wave is so named because it has sharp, or "square," edges. If the luminance of the square wave was plotted as a function of spatial position, it would look like the function shown at the top of Figure 4.10. According to Fourier's theorem, the square wave is composed of a sine wave of the same spatial frequency, called the fundamental frequency, plus a set of sine waves that form a series that are odd multiples of the frequency and amplitude of the fundamental. The latter waves are called harmonics. In the case of the square wave, these harmonics include sine waves that are three times the fundamental frequency and one-third the amplitude, five times the fundamental frequency and one-fifth the amplitude, seven times the fundamental frequency and one-seventh the amplitude and so on to infinity. Figure 4.10 shows how the addition of each successive harmonic component makes the combined sine

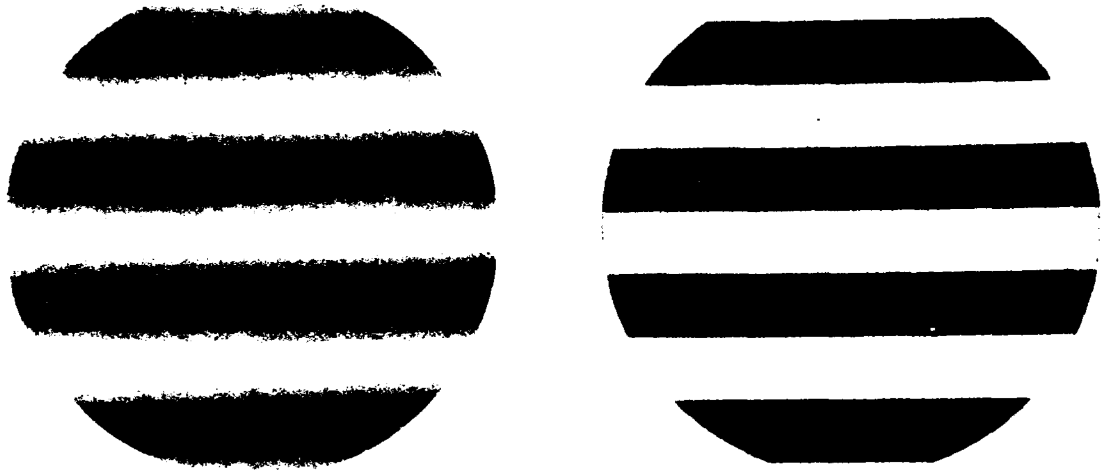


Figure 4.9. Sine-wave (left) and square-wave (right) gratings of the same spatial frequency. (from De Valois & De Valois, 1988)

waves look more and more like a square wave. While a mathematically perfect square wave requires an infinite number of sine waves, only the frequencies to which we are sensitive (as defined by the contrast sensitivity function) need be used. This can be demonstrated by producing a set of sine waves and adding various components until the complex wave becomes indiscriminable from a true square wave.

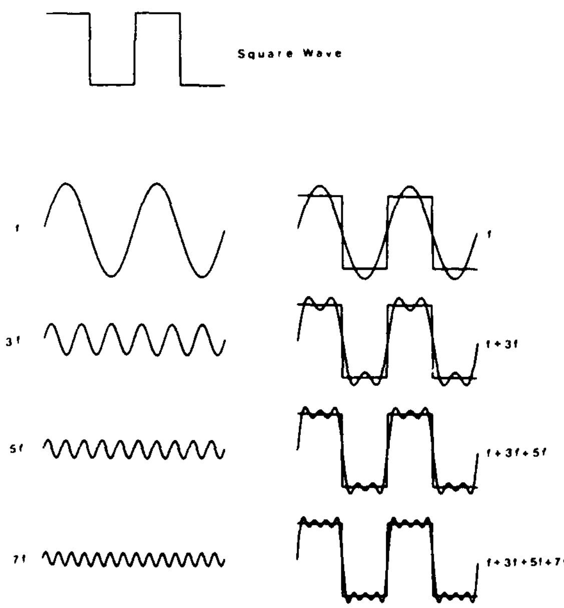


Figure 4.10. Illustration of Fourier synthesis of a square-wave (top left) and waveform changes as various sinusoidal components are added. (from DeValois & DeValois, 1988)

Our demonstration of Fourier synthesis of a square wave involved energy variations only along one dimension, i.e., a vertical grating only changes from left to right. To synthesize natural images from a set of sine waves, one must add the sinusoidal energy variations in two dimensions. Figure 4.11 shows how a set of sine waves can be progressively summed in two dimensions to produce a complex pattern. The top left shows the fundamental frequency and to the immediate right is the power spectrum -- a two-dimensional graph of the frequency, amplitude, and orientation of the sine wave components. Each successive frame shows the number of spatial frequency components in the picture. Although the computer screen generated the image using about 65,000 points, the picture is recognizable with only about 164 spatial frequencies.

Fourier analysis has been used in psychophysical experiments to successfully predict performance on visual detection, discrimination, and recognition tasks with complex stimuli (for reviews see Sekuler, 1974; DeValois & DeValois, 1988). This approach involves a number of assumptions that are true only under a restricted set of conditions. The advantages of this approach, however, should be obvious for evaluating displays. Under some conditions, the contrast sensitivity function might be used as a filter through which the visibility of

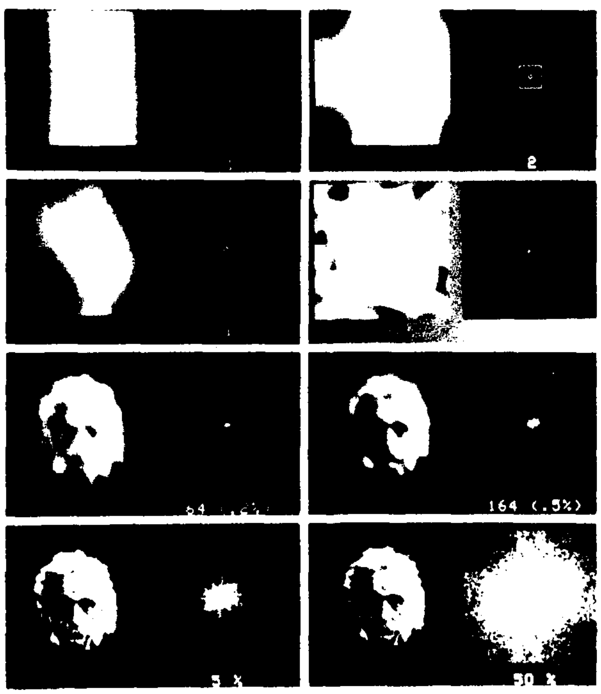


Figure 4.11. Illustration of Fourier synthesis of a complex image by the successive addition of sinusoidal components in two dimensions. (from DeValois & DeValois, 1988)

various components of a pattern or the whole pattern can be predicted. Studies conducted at the Boeing company have used the contrast sensitivity function to predict image quality. Since the contrast sensitivity function defines the energy required for a threshold response, the energy above threshold should contribute to pattern visibility. When this suprathreshold energy is summed across spatial frequencies, it correlates highly with subjective measures of image quality (Klingberg et al., 1970). The contrast sensitivity function has also been used with success to predict image quality of other display parameters such as target size, display background, and clutter (Snyder, 1988).

Form-Color Interactions

If the eye is alternately exposed to a red vertical grating and a green horizontal grating for about 5-10 minutes while the observer freely scans, there will be a powerful aftereffect. If a black-and-white grating is used as a test stimulus, following adaptation the observer will see the white region as green when the stripes are vertical and as red when they are horizontal. These aftereffects are in the opposite direction to the adapting condition and are contingent on the orientation of the test pattern. This is known as the McCollough (1965) effect. Color-contingent aftereffects under these conditions are quite long-lasting -- up to months in some cases -- and cannot be attributed to traditional after-images. Effects of this sort are not uncommon for individuals who work on video display units. Exposure to red or green symbology on a display with a dark background would later be expected to cause white letters to appear green or red, respectively.

Depth Perception

Information about size, color, contrast, and motion are not all that we need to understand our visual environment. We also need to perceive the positions of objects in space, an ability called depth perception. There are two major classes of cues that we use to perceive depth. ***Monocular depth cues*** provide information about depth that can be extracted using only one eye. ***Binocular depth cues*** rely on an analysis of slightly different information available from each of the two eyes.

Monocular Depth Cues

If you close one eye and look around, you will probably not be confused about the relative distances of most objects. Your perception of distance in this case is based on monocular cues which are even more powerful than some of the binocular cues to depth (Kaufman, 1974).

The *size* of objects can sometimes indicate their relative depth. If several similar items are presented together, the larger items will be judged as closer. For example, the series of circles in Figure 4.12 appears to be receding into the distance. This makes sense because, in fact, the size of an object's image on the retina becomes progressively smaller as it moves away.

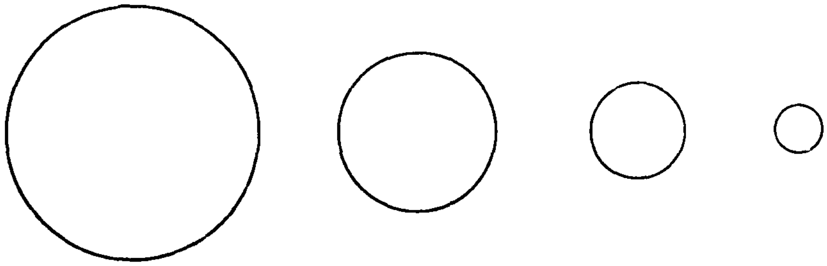


Figure 4.12. Illustration showing how the size of an object influences the perception of distance.

The ability to infer distance from image size often depends on familiarity with the true size of the objects. At great distances, such as looking down from an airplane, we perceive objects to be smaller than when they are near. In this situation, our familiarity with objects and their constancy of size serve as a source of information about distance. Although from the air a house seems like a toy, our knowledge about the actual size of houses informs us that the house is only farther away, not smaller.

The relation between size and distance can lead not only to faulty inferences about distance, as illustrated by Figure 4.12, but assumptions about distance can also lead to faulty inferences about size. When we are misinformed about distance, our perceptions of size and shape will be affected. You have probably noticed, for example, how much larger the moon appears when it is low on the horizon than high in the evening sky. This is called the ***moon illusion***. The change in the moon's appearance is only slightly affected by atmospheric phenomena; by far the greatest effect is perceptual. Our retinal image of the moon is the same size in both positions. You can prove this by holding at arms length a piece of cardboard just large enough to block the moon from view. The same piece of cardboard blocks the moon at the horizon and at its zenith equally. Though they look different, they measure the same. The moon illusion seems to be caused by inaccurate distance information about very far objects (Kaufman & Rock, 1962). Because we see intervening objects on the earth's surface when we look at the moon near the horizon, our internal distance analyzers apparently cue us that the moon is farther away than when it is at its zenith. An object analyzed as more distant has to be larger to produce an image

of the same size. Thus we perceive the moon as larger on the horizon than when it is at its zenith.

The relationship between size and distance is important to understanding not only harmless illusions such as the size of the moon, but also in situations of more significance. As mentioned above, changing fixation from a head-up display to distant objects often requires a change in the state of accommodation. Change in the focus of the eye is accompanied by a change in the apparent visual angle of distant objects. Thus, when a pilot shifts fixation from a HUD to a distant surface in the outside world, the objects in the distance may appear smaller and more distant than they really are (Iavecchia, Iavecchia & Roscoe, 1988). While the resultant spatial errors in perception are temporary, Iavecchia et al. believe it could introduce a significant safety hazard under some conditions.

Any ambiguity about relative distance in relation to size can be rectified when one object partially occludes another, as shown in Figure 4.13. We perceive the partially occluded object as being more distant. This cue to depth is called ***interposition***.

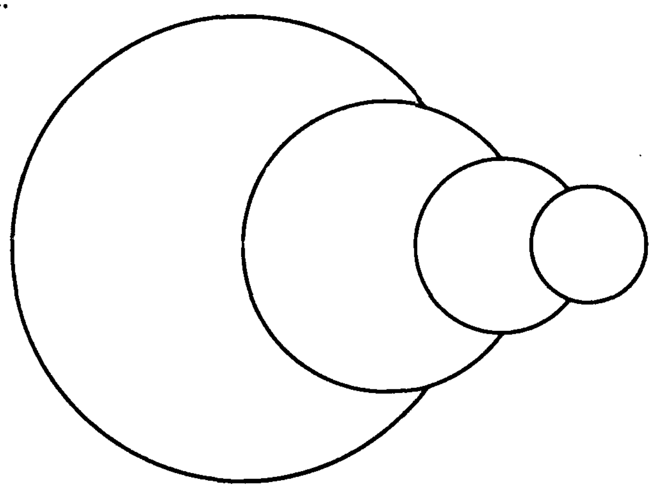


Figure 4.13 Illustration of interposition as a monocular cue for distance.

If a distant object is not partially occluded, we may still be able to judge its distance using ***linear perspective***. When you look at a set of parallel lines, such as railroad tracks going off into the distance, the retinal images of these lines converge because the visual angle formed by two points parallel to another decreases as the points are farther away. This cue to depth is so powerful that it may cause objects of the same size to be perceived as different, as in Figure 4.14.

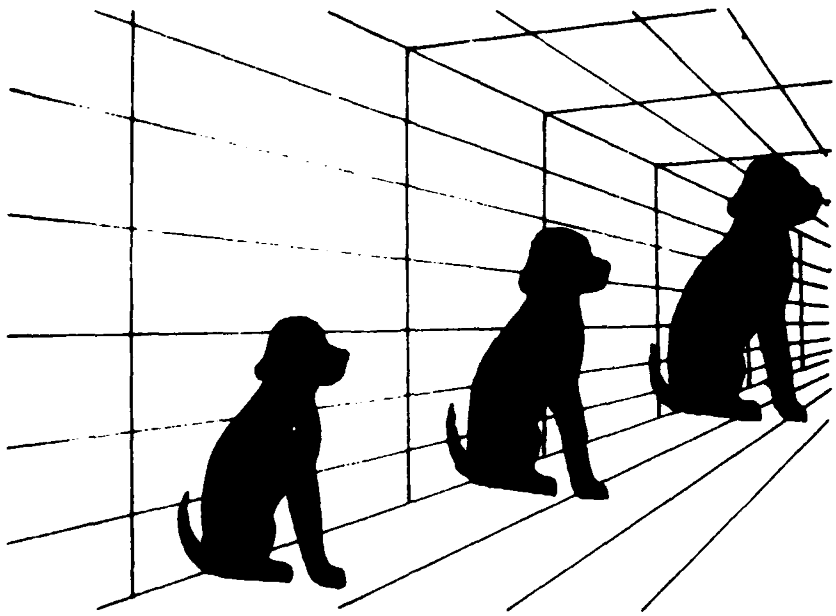


Figure 4.14 Illustration of how linear perspective makes the same size objects appear to be different sizes. (from Sekuler & Blake, 1985)

If you look at a textured surface such as a lawn, two blades of grass the same distance apart would be separated by a smaller distance in the retinal image the f·rther away they are because they cover a smaller visual angle. Most surfaces have a certain pattern, grain, or *texture* such as pebbles on the beach or the grain of a wood floor. Whatever the texture, it becomes denser with distance. This information can provide clear indications of distances (Newman, Whinham & MacRae, 1973). Figure 4.15 shows how discontinuities in the texture also indicate a change such as an edge or corner.

Of special relevance to aircraft pilots is the depth cue known as ***aerial perspective***. As light travels through the atmosphere, it is scattered by molecules in the air such as dust and water. The images of more distant objects are thus less clear. Under different atmospheric conditions, the perceived distance of an object of fixed size may vary. For example, an airport will appear farther away on a hazy day than on a clear day.

Some monocular cues to depth are not static, but are dependent on relative movement. When we are moving, objects appear to move relative to the point

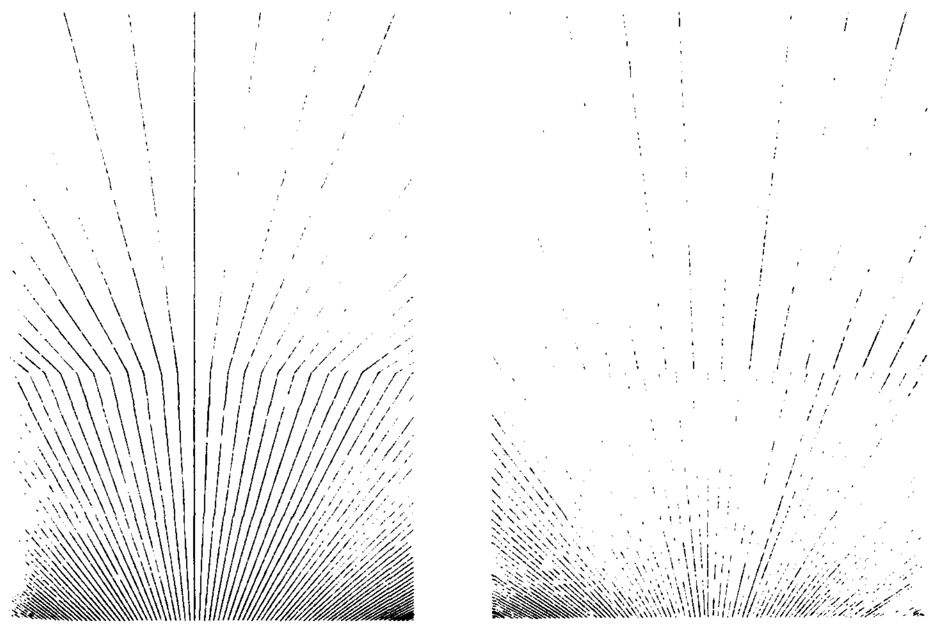


Figure 4.15 Illustration of texture gradients as a cue to distance. (from Gibson, 1966)

of fixation. The direction and speed of movement is related to their relative distances. This is illustrated by Figure 4.16. Objects that are more distant than the point of fixation move in the same direction as the observer. Objects in front of the point of fixation move opposite to the direction of the observer. You can demonstrate this by holding two fingers in front of you at different distances and then observing their relative displacement as you move your head back and forth. The difference in how near and far objects move, called **motion parallax**, is probably our most important monocular source of information about distance. Motion parallax occurs from any relative motion -- moving the whole body, the head, or the eyes.

Motion perspective is a phenomenon related to motion parallax. It refers to the fact that as we move straight ahead, the images of objects surrounding the point of fixation tend to flow away from that point. Figure 4.17 illustrates motion perspective for an individual walking through the stacks of books in a library. If the observer were to back up, the flow pattern would contract rather than expand. These optic flow patterns carry information about direction, distance and speed, and are believed to be an important depth cue used by pilots to land planes (Regan, Beverly & Cynader, 1979).

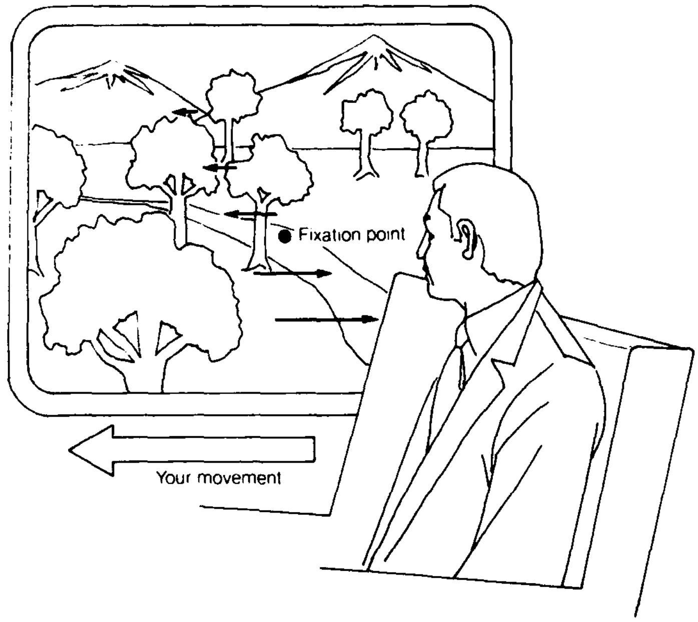


Figure 4.16. Illustration of motion parallax. (from Coren, Porac, & Ward, 1984)

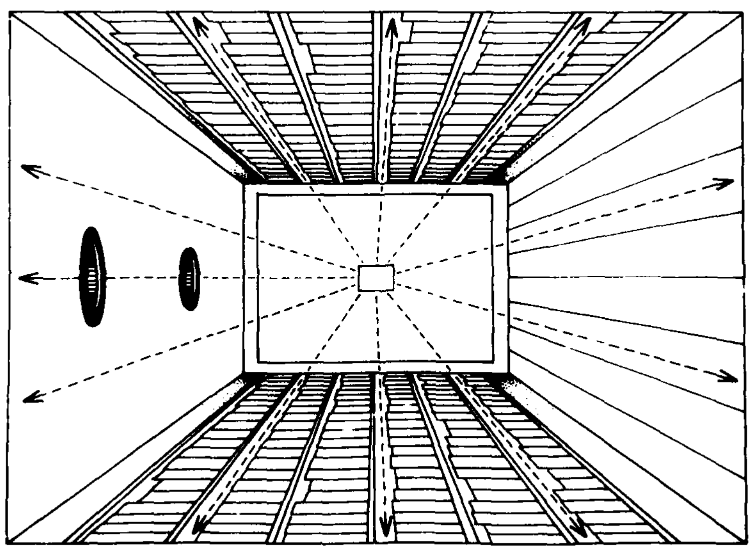


Figure 4.17. Illustration of motion perspective for a person who is moving and fixating straight ahead. (from Matlin, 1983)

Ocular Convergence

When fixating a distant object, the image of the object will fall on the fovea of each eye. As the object is brought nearer, maintenance of fixation will require that the two eyes move inward or converge. This information about convergence can be used to gauge the absolute distance of objects, provided the objects are not more than about 10 feet away. Beyond this distance, the convergence angle of the two eyes approaches zero.

Stereopsis

Because the two eyes are separated by about 3 inches, the visual fields are slightly different for the two eyes (refer back to Figure 2.16 in chapter 2). In the region where the two eyes have overlapping visual fields, they will receive slightly different images of objects. This is easily verified by alternately fixating an object a few feet away with one eye and then the other. With the left eye you will see more of the left side of the object, and with the right eye you will see more of the right side of the object. This difference between the images in the two eyes is referred to as ***retinal disparity*** or ***binocular disparity***.

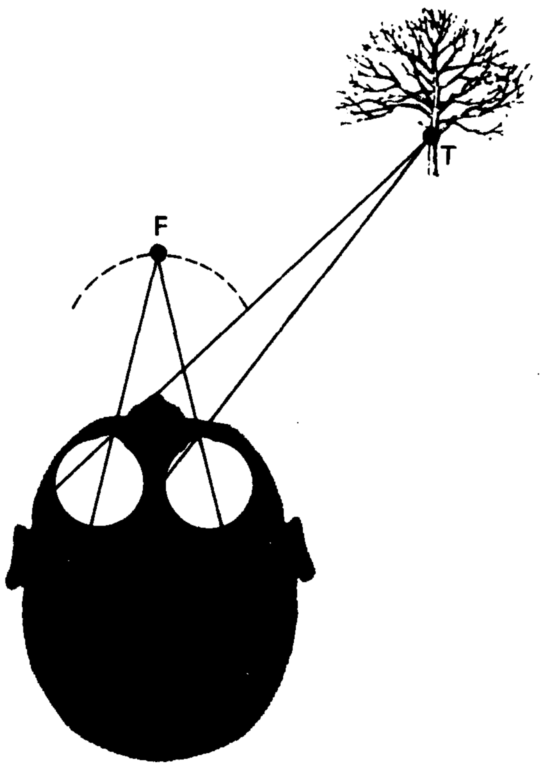


Figure 4.18. Schematic illustration of binocular disparity. (from Werner & Schlesinger, 1991)

Figure 4.18 shows how binocular disparity arises. When we fixate on point F, both eyes are oriented so that the image falls on the center of the fovea in each eye. Images from objects at other distances from our eyes -- for example, the tree in Figure 4.18 -- will fall onto different locations in relationship to the foveas. This happens because the two eyes have different angles of view. Images of objects that are either inside or outside the half-circle in Figure 4.18 will strike the two retinas differently. Thus, disparate signals from each eye will be sent to the brain where comparisons are made by specialized cells; different cells are tuned to respond

according to the amount of disparity (Pettigrew, 1972). The amount of binocular disparity (specified in arc units) provides us with information about how far in front or behind our fixation (F) point an object is. The ability to judge depth using retinal disparity is known as ***stereopsis***.

There are several ways to demonstrate stereoscopic depth from two-dimensional images. Wheatstone (1838) showed that if one image is presented to one eye and another image to the other eye through a stereoscope, the images could be fused and a three-dimensional image could be seen. Today, 3-D movies are created by projecting two (disparate) images on a screen. Separation of the images is made possible by projecting them with polarized light of orthogonal orientations. If the viewer has polarizing glasses, the two images will be separately projected to each retina, fused, and perceived as three dimensional.

The remarkable ability of the brain to extract information about depth was demonstrated by Julesz (1971) through patterns called random-dot stereograms. A random-dot stereogram is shown in Figure 4.19. The two squares consist of dots placed randomly within the frame. However, in one frame the dots from a small square region were displaced (moved) slightly. When these two images are presented separately to each eye the displaced dots will produce retinal

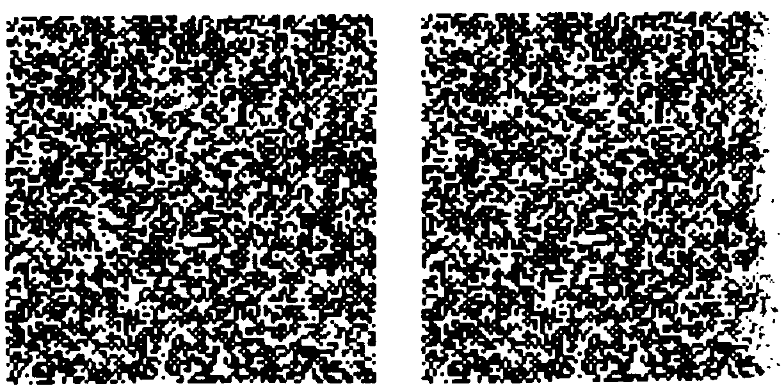


Figure 4.19. A random-dot stereogram (from Julesz, 1971)

disparity. Thus, we will perceive the subset of dots within the square as lying in front or behind the other dots. The ability of the visual system to correlate all of these random dots shows that retinal disparity does not require a comparison of specific forms or features of objects. One possible basis for extracting the information in the two eyes quickly might be for the visual system to process the spatial frequency content in the two images (Frisby & Mayhew, 1976).

Random-dot stereograms demonstrate the keen sensitivity of the human visual system to binocular disparity. There are many practical implications of this ability. For example, if a counterfeit dollar bill is placed on one side of a stereoscopic viewer and a genuine dollar bill on the other, the two can be compared and differences of 0.005 mm can be detected because they will stand out in depth. Other virtues of stereovision are well known to aerial surveyors and experts in aerial surveillance. Under optimal conditions, stereoscopic depth can be used to resolve displacement in depth of about 2 sec of arc. This corresponds to a difference that is smaller than the diameter of a single cone receptor.

Stereoacuity varies with the distance of the object. Beyond about 100 feet, retinal disparity diminishes so greatly that this cue to depth is not useful. Thus, it is sometimes noted that routine aspects of flying an airplane do not require stereopsis, but it is helpful when moving the plane into the hangar (DeHaan, 1982).

Binocular Rivalry

If the scenes presented to each eye are very different, such as when the images of objects are too binocularly disparate, the visual system does not fuse the images. Rather, views of the two scenes may alternate from one eye to the other or a mosaic that combines portions of the two images may alternate. This is known as ***binocular rivalry*** and can occur whenever the images presented to each eye are too different to be combined. Apparently the visual system attempts to match the images from the two eyes and when this cannot be done, one of the images or at least portions of one image are suppressed.
During early life, the images to the two eyes may be chronically discordant due to the two eyes being improperly aligned, a condition known as ***strabismus***. If this condition is not corrected in early childhood, the input from one of the eyes may become permanently suppressed and the individual will be ***stereoblind***, that is, incapable of using stereoscopic cues to depth. Whether due to strabismus or other causes, about 5-10% of the population is stereoblind (Richards, 1970).

Color Stereopsis

When deeply saturated colors are viewed on a display, it sometimes appears that the different colors lie at different depths. This phenomenon, known as color stereopsis or chromostereopsis, is illustrated in Figure 4.20. The effect is most clearly seen with colors that are maximally separated in the spectrum. On displays, red may appear to be nearer than blue.

Figure 4.20. An illustration of color stereopsis.

Color stereopsis is due to retinal disparity arising from chromatic dispersion by the optics of the eye. Short wavelengths are imaged more nasally than long wavelengths and the resultant retinal disparity leads to the perception that the different colors are at different depth planes. As pointed out by Walraven (1985), display operators can minimize this effect by using less saturated colors or brighter backgrounds.

Implications for Displays

Stereopsis provides a little used channel for presenting information on visual displays. By using retinally disparate images, it is possible to create more realistic portrayals of the external environment than would be possible on displays carrying only monocular information. Whether stereo imagery on displays would improve performance in the cockpit should be further studied. There is some evidence that it can decrease response time, increase recognition, and reduce workload (Tolin, 1987).

Perhaps the most interesting applications of stereo displays are not in the cockpit, but in the control tower (Williams & Garcia, 1989). The workload of traffic controllers could conceivably be reduced if aircraft could be seen in three-dimensional rather than two-dimensional space. Methods for generating such "volumetric" displays and evaluation of human performance with these displays provide an interesting challenge for the future.

Chapter 5

Information Processing

by Kim M. Cardosi, Ph.D., Volpe Center
based on material presented by Peter D. Eimas, Ph.D., Brown University

What Is the Mind?

An important belief shared by cognitive psychologists is that the mind has many components that perform different functions. We can measure the time it takes for the different parts of the mind to do their jobs, even though our experience of information processing or of any cognitive function is that it happens instantaneously. In laboratory research, psychologists can parcel our mental processes into component parts and measure the time it takes for each component task to be accomplished.

The Brain as an Information Processor

Figure 5.1 shows one representation of the mind as an information processing system. This system is a product of the brain. There are at least 10 billion, probably 100 billion cells called neurons in the brain. Each neuron has between a hundred and 10,000 connections. It is a very large system and its size which permits us to perform the many mental tasks that we do so well, for example, communicate by means of language, solve problems, and monitor complex physical systems that inform us about events in the environment.

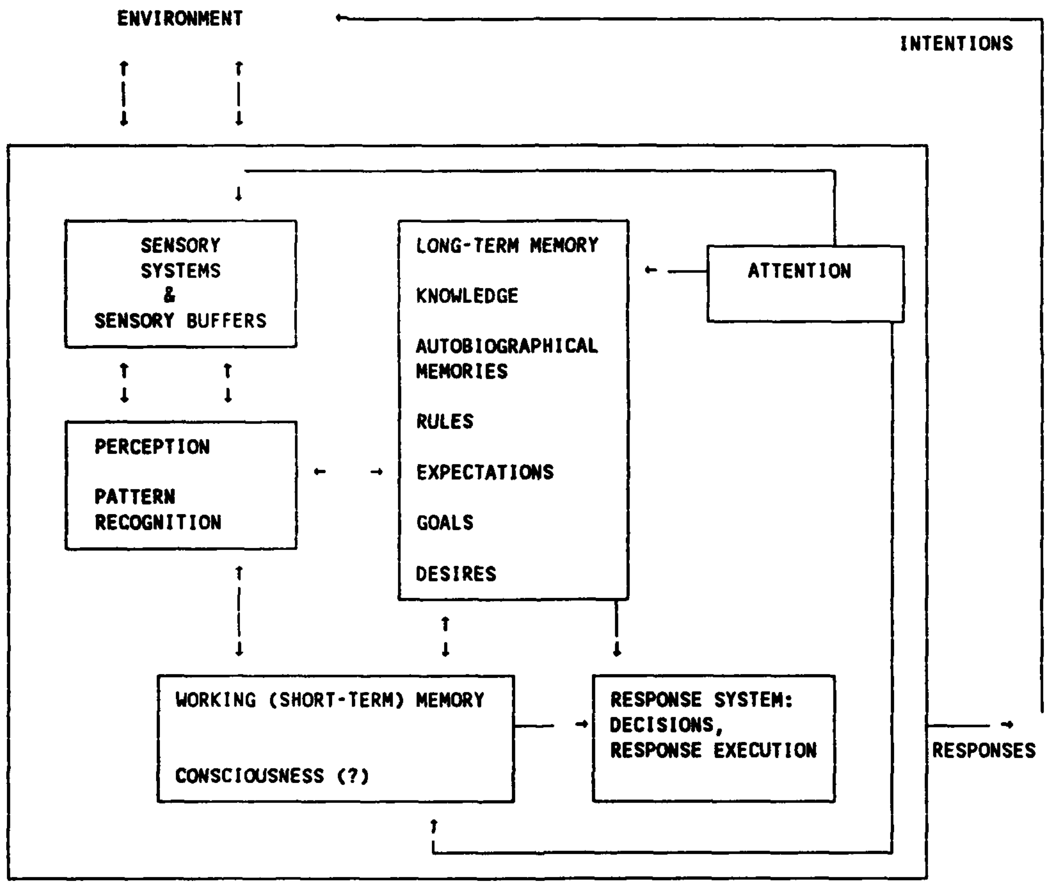


Figure 5.1 Boxology diagram of mental processing. (original figure)

Information from the environment comes in through sensory systems: both internal and external. An example of an internal sense is hunger pangs. External senses are sight, hearing, smell, taste, and touch. Both internal and external senses provide our minds with information that flows through the system, and results in a response. Our responses or behavior can change the environment and create a new situation. Then we may respond in a different way to the new situation we have helped to create. That's why the diagram shows the arrows going from response back to the environment.

The first box in Figure 5.1 represents the senses, which actually detect the things and events in the environment, and the sensory register. In the sensory register, information is held for a very brief period of time (less than one second) while it is selected or filtered and ultimately processed so as to provide us with the percepts that we experience. Information is processed by means of what has been called a ***pattern recognition*** system. We have patterns (such as your name, a familiar voice, an aircraft call sign) stored in ***long-term memory*** that help us with the recognition process. Long-term memory is the memory you have for your entire life. This includes all of the knowledge you have, what you've learned in school through all the years, the expertise you've gained in your work, etc. It also has your autobiographical memories, what you did when, with whom. To recognize a pattern, to know what something is and its significance, means you have matched it to something you already know.

When we recognize selected information, we hold it in ***short-term memory***, also called ***working memory***. Short-term memory is like the central processing unit of a computer. It's where we do our work, where we solve our problems -- at least partially -- where we bring information together from short- and long-term memory that begins to answer the questions that are posed to us by our environment. Short-term memory has a limited capacity. In it, we can store approximately five to nine items (e.g., letters) or chunks of information (e.g., words) for up to one minute. Information that can be retrieved after one minute is said to have been transferred to long-term memory. Long-term memory has unlimited capacity, but retrieval can be a problem. That is, the information is known to be in long-term memory but it, at least temporarily, cannot be transferred to short-term memory for use. Memory will be discussed in more detail later in this chapter.

In summary, information can be viewed as constantly moving back and forth between the outer world and the mind through our internal and external senses. The information is filtered, processed by pattern recognition systems and stored briefly in short-term memory, which may also be the site of consciousness, and can under the right circumstances, be stored in long-term memory indefinitely. This information in working memory can also be used to

make a decision and initiate a response. Decision making and response selection will be covered in Chapter 7.

We can classify our stored knowledge as explicit or implicit. ***Explicit knowledge*** is knowledge that you have direct and immediate access to. This includes your name, your phone number, what you do for a living, who your spouse is, what your children's names are, all the knowledge you have about your expertise and your profession, etc. All of these are explicit forms of knowledge that you can describe in detail as well as use for many tasks of a cognitive nature.

Implicit knowledge is knowledge that you have, but you are not able to describe; that is to say, you do not have direct access to this knowledge. Good examples of implicit knowledge are things like riding a bicycle, playing tennis, catching a baseball, the syntactic rules of your language, etc. Most likely, unless you're a physicist, you have no explicit knowledge of the laws of physics that you use when doing such things as riding a bicycle. Nevertheless, you can do them properly. Your implicit knowledge enables you to do so -- it is available for certain tasks, but it is not available to consciousness.

Figure 5.1 breaks things up rather neatly, as if these processes occur separately, taking a lot of time. However, information processing, perception, speaking, and listening go on very, very quickly. The diagram shows mental activity occurring in accord with a ***serial processing*** system; that is, we do one thing at a time. However, there is the belief that ***parallel processing*** (doing more than one thing at a time) also occurs. It is difficult to substantiate that parallel processing goes on in the mind because the measuring instruments are limited. We can measure the electrical activity of someone's brain and say that the brain is working because we see the blips on the electroencephalogram. We can be much more precise and say that certain areas are working. What appears to be true is that some of those areas are working in parallel. Indeed, if we think of all the events that must occur during perception of visual scenes or spoken language, parallel processing would seem to be absolutely necessary if we are to explain how these processes, these mental activities, could occur so quickly.

Another box in Figure 5.1 is ***attention***. Attention is simply the part of the mental system that directs us to one sort of information rather than another. We are able to attend to a particular stimulus even in the presence of an enormous amount of other stimulation. This ability to selectively attend to specific information will be discussed in detail in Chapter 8.

Information processing takes time, as noted above. The time required to process information depends upon many factors. In most cases, information will be

processed only to the extent that is required by the task. The more complex the task, the more time will be required. For example, in any array of colored numbers, more time will be needed to count the blue numbers than to decide whether or not blue numbers are present in the display. Still more time will be needed to add the blue numbers. This type of difference in the level of processing is referred to as "depth" of processing. The more, or "deeper," the information is processed, the easier it will be to remember (Craik and Lockhart, 1972). For example, a controller is more likely to remember "seeing" an aircraft that he or she has communicated with several times than one with which no communication was required. In our previous example of an array of colored numbers, the person who added the blue numbers would have more success in recalling them than the person who counted the same numbers. Information that is not specifically attended to is not likely to be remembered. The more attentional resources spent on processing the information, the more accurately the information will be remembered. This has implications for complex tasks in which it is important to remember certain pieces of information. We can maximize the chances of being able to remember information by requiring that the information be used or processed in some way. Information that is not actively attended to will not be easily recalled from memory when needed.

Attention

In *Principles of Psychology* (1890), William James defined ***attention*** as "the mind taking possession, in clear and vivid form, of one of what seems several simultaneous possible objects or trains of thought. It implies withdrawal from some things to deal effectively with others." In processing information we can focus on specific information at the expense of other information, and we can shift our attention from one thing to another. What are the costs of focused attention? How do you move attention around?

Attention directs us to something particular. Some researchers consider human mental processing to be, for the most part, a serial processing system like the central processing system of most computers. Computers do one thing at a time, but they do them very, very quickly; performing millions of operations per second. Our neurons are not as fast. In fact, they're incredibly slow. So what we probably do is group great masses of them together to do things and use parallel processing. One mass of neurons in one section of the brain does one thing, while another mass in another section does another thing.

The attention mechanism that directs our processing energy works both within a sensory modality, (i.e., within vision or within audition) and across modalities. There may be two types of attention: one that directs you to a modality, and one that works within a modality. Alternatively, there may be a

single central processor in the mind that is responsible for prioritizing incoming information.

Selective Attention

Some of the early scientific work on attention began around 1950 in a group led by Donald Broadbent in England. It took its impetus from a phenomenon that came to be called "the cocktail party effect." If you go to a cocktail party where there are only a couple people and the noise level is not bad, it's easy to understand the person you're talking to. After 150 people have arrived, the noise level is overpowering; if you recorded it, it would sound like gibberish. It would be very difficult to pick out one conversation on the tape and pay attention to it. However, an individual at the cocktail party can begin to and continue to attend to a speaker and understand what that speaker is saying despite distracting noise.

One factor that makes this possible is the distinctiveness of the voice of the person who is speaking to you; it is easier to attend to an individual if the voice is distinctive in some way. For example, it would be easier to attend to a woman's voice when the distracting voices were men's voices, because of the pitch of a woman's voice tends to be very different from a man's. Another factor is the direction of the voice. You can focus on a voice by virtue of the direction it comes from: a voice coming from a certain direction hits one ear earlier than the other by a very precise amount of time. Other factors that allow you to attend to a particular voice at a noisy cocktail party are the coherence or meaning of the speech, the nature of the voice, and the emphasis given to the words. These kinds of simple matters were related by Broadbent to what was called "picking up a channel of information and staying attached to it."

Neisser and Becklen (1975) performed interesting experiments that show the power of selective attention. They showed videotapes of games to subjects and had them perform simple tasks. In one tape, three men bounced a basketball back and forth to each other. The subjects' task was to count the bounces. Then, Neisser showed a tape of two people playing a handslapping game. The subjects' task here was to count the number of hits. If either task was performed alone, counting accuracy was near perfect. When the two tapes were superimposed, it was still quite easy to count either the number of ball bounces or the number of hand slaps. Trying to count both at the same time, however, was quite difficult. It was so difficult that the subjects failed to notice the "odd" events of the ball disappearing or the men being replaced by women. This is one example of the filtering of information. We can attend to and process

complex information quite efficiently. However, if the task is attentionally taxing, we may not process all of the information available to us.

Filtering can occur at both high and low levels. In low-level filtering, also called ***early selection***, the person can respond to stimuli more quickly because simpler processing (e.g., male vs. female voice) allows him or her to decide which information is pertinent. High-level filtering, also called ***late selection***, demands more effort because you have to process the meaning of something, not just the simple, physical characteristics of it. In this case, it is more difficult for the person to filter out the unimportant information and decide what is pertinent. Whether your selection of pertinent information occurs early or late depends upon the task.

The Cost of Multiple Tasks

Johnston and Heinz (1978) sat people in front of a display box and instructed them to press a button whenever a light came on. The light came on at random intervals. Subjects simultaneously listened to a tape of excerpts from Reader's Digest articles. Their task was to listen to the tape and press the button when the light came on. The participants also had to answer simple true/false questions about the passage at the end of each trial. These questions were asked to ensure that the subjects attended to the tape and didn't neglect the button-pressing task. Adding the task of listening to a message raised the time required to respond to the light from 320 msec to 355 msec. Thus, there was a small, but statistically significant, rise in response time for a very simple task (i.e., a button press) when another simple and unrelated task (i.e., listening) was added to it. As the experimenters made the listening task more difficult (e.g., attend to one of two stories), response time rose with the difficulty of the task. For example, it took an average of 387 msec to press the button in response to the light as subjects tried to pay attention to one of two very different messages (i.e., on different topics with one spoken by a man and one spoken by a woman), and an average of 429 msec to respond to the light as subjects tried to attend to one of two very similar messages (i.e., with same sex speakers and similar content).

These experiments demonstrate three things. First, the time required to conduct even the simplest task will increase as other, even simple and unrelated tasks, are added to it. Second, the more difficult the added task is, the higher the attentional cost due to the additional burden on the attentional mechanism. Third, this attentional cost can be measured in the laboratory. On the average for these subjects, it took 320 milliseconds to simply press the button when the light came on without any information being broadcast to the ears. If a stimulus (e.g., a warning light or text message) appears directly in front of a person,

response time to it will be faster than if eye movements are required to fixate, or focus on, the information. Similarly, if the stimulus appears within the person's visual field, but in the periphery rather than at the fixation point, response time will be lower than when an eye movement is required, but higher than when the target appears at the fixation point. While we usually move our eyes when we shift attention, this is not always necessary. We can shift our mental focus, or ***internal attention***. Even when shifting internal attention does not involve eye movements, it does take time. The time required to shift internal attention increases with the distance from the fixation point and travels at a velocity of about 1° per 8 msec. (Tsal, 1983). Furthermore, some information that is presented during the time that it takes to make this shift may not be processed (Reeves and Sperling, 1986).

Automatic and Controlled Processing

Automatic processing occurs in highly practiced activities like driving a car, riding a bike, etc. You do it without necessarily being aware of what you're doing. It just happens. Automatic processing is fast. It appears to be parallel, that is, you can do more than one thing at a time, and it's fairly effortless. Controlled processing means voluntary, one-step-at-a-time processing. It is a rather slow process. It requires focussing attention to specific parts of complex tasks. Acquiring controlled processing can be done simply by saying "Pay attention to this." Acquiring automatic processing, on the other hand, may be very slow or fast, depending on the task. At very low levels where the distinctions are being made by simple kinds of physical stimuli, e.g., search for a red object among varying colored objects, search for a curved line among all straight lines, automatic processing can be achieved quickly. Things tend to jump out. It's called the ***popout effect***. If you ask subjects, "How did you find the red square?" they say, "Well, it was kind of there. It popped out at me." Whether there was one choice, two choices, or four choices, they were trying to search for, it really didn't make any difference. It just seemed to show up to them. They had to do less processing. It popped out. Something was "automatically" happening to them. If you have to do high-level processing, such as searching for particular letters in a field of other letters over and over again, achieving automatic processing is much more difficult and takes much more time. Automatic processing allows for development of fast, highly skilled behaviors without eating up attentional resources.

Many things we do acquire a quality of automaticity, which is to say we do these things automatically, without thinking much about them. For example, learning to drive a car is a complex, difficult task. It is attentionally taxing and even simple conversation is very distracting. An experienced driver, however,

can drive and carry on a conversation with ease. This is what we mean by automatic processing. You can perform your primary task (e.g., driving) and simultaneously perform another task (e.g., conversing) and do each as well as if you were doing it alone. And, you are doing both without a great deal of stress and effort because one of these tasks is being done automatically.

There are many examples of complex, difficult tasks becoming easier and less taxing with practice. Any difficult task is, at first, attentionally all consuming; extraneous or unexpected information is not likely to be processed. Sufficient practice, however, can make even the most difficult tasks sufficiently easy to deal with other incoming information. This is the advantage of automaticity. When tasks or parts of tasks (subtasks), such as flying straight and level, are performed automatically, resources are available to perform other tasks simultaneously. While it is easy to see the benefits of automaticity, it is important to be aware of the hidden costs. One of these costs is commonly referred to as complacency. Since we devote less attention to tasks we can perform automatically, it is easy to miss some incoming information - even when this information is important (such as a subtle course deviation or a new stop sign on a road traveled daily). We are most likely to miss or misinterpret information when what we expect to see or hear differs only slightly from what is actually there.

Expectation

Expectations are powerful shapers or perception. We are susceptible - particularly under high workload - to seeing what we expect to see and hearing what we expect to hear. Even when we do notice the difference between the expected and the actual message, there is a price to pay; it takes much longer to process the correct message when another one is expected than when the correct one is expected or when there are no expectations.

Scharf, Quigley, Aoki, Peachey, and Neeves (1987) demonstrated that even the simplest of information processing shows a detrimental effect of a discrepancy between the expected and actual information. They played a pure tone between 600 and 1500 Hz that was just barely audible and told subjects that this tone would be played again during one of two time intervals. No tone was played in the other interval. The subjects' task was to decide in which interval the tone was played. When the tone that the subjects had to listen for (the target) was the same frequency as the one they had heard first (the prime), subjects were 90% correct in identifying the interval that contained the tone. When the frequency of the target was changed, performance suffered. For example, when a 600 Hz tone was expected and a 600 Hz tone was the target, performance was near perfect with 90% accuracy. When a 1000 Hz tone was expected and a

600 Hz tone was the target, performance was near chance with subjects guessing which interval contained the tone with only 55% accuracy. The same was true when the target tone was 1500 Hz and the prime was 1000 Hz. Even a difference of only 75 Hz (with targets of 925 and 1075 Hz) resulted in a drop in accuracy from 90% to 64%. This supports transfer of training principles which will be discussed in detail in Chapter 7. The closer auditory warnings are to what is expected (e.g., from simulator training or experience in other aircraft), the easier it will be to "hear," all other things being equal.

The powers of expectancy are even more obvious in higher level processing, such as speech perception. If you quickly read aloud, "the man went to a restaurant for dinner and ordered state and potatoes," chances are any listeners would hear "the man went to a restaurant for dinner and ordered steak and potatoes." It is not surprising that there have been many ASRS reports of pilots accepting clearances not intended for them after requesting higher or lower altitudes. Again, we are most likely to make such mistakes when what we expect to hear is only slightly different from what should be heard (as with similar call signs).

Pattern Recognition

Pattern recognition is one of the components of our model of information processing (Figure 5.1). The word "pattern" refers to anything we see or hear or really sense by any means. Our ability to perceive and identify patterns - whether words or objects - depends heavily on our ability to match the pattern that we see or hear with the representations of patterns that are stored in memory. We refer to this matching as pattern recognition.

There have been many theories of pattern recognition. The ***template theory*** states that there are entire patterns stored in our brains as whole patterns. When we see or hear something, we match this to one of the stored patterns to identify it. The problem with this theory is that we would need an infinite number of templates to match the innumerable ways in which an object may be presented to us -- one stored pattern for each different pattern in a different size and orientation. For example, consider an individual letter "Z." This letter may be presented to us in print (in either upper or lower case) or handwritten by many different writers. While no one template would fit all of these Z's, we usually have no trouble recognizing Z's as such.

A similar theory, the ***feature theory*** states that incoming information is broken down into its component physical characteristics or features and their relations. A "Z" for example, can be broken down into two horizontal parallel lines, an

oblique line, and two acute angles. There is some physiological evidence to suggest that our brains do process some information in this way. There are brain cells that respond only to horizontal lines, others that respond only to vertical lines, etc. But that does not mean that we process all information in this way. In fact, it would be difficult to explain the identification of most real world objects in this way. For example, by what features do we recognize a dog? There are barkless dogs, tailless dogs, dogs with three legs, hairless dogs, etc. Whatever feature we might consider using to define "dog," we are sure to think of an exception.

The template and feature theories both assume a "bottom-up" mode of information processing. That is, they assert that we process information by beginning with the physical aspects of the stimulus and working up to its meaning. In a "top-down" approach, the meaning is accessed first or at least in parallel with other information (usually with the aid of contextual cues), and then that information helps us process the physical features. For example, none of the characters in Figure 5.2(a) appear at all ambiguous; there is a clearly definable "A", "B", "C", and "D". However, in a different context. the "B"

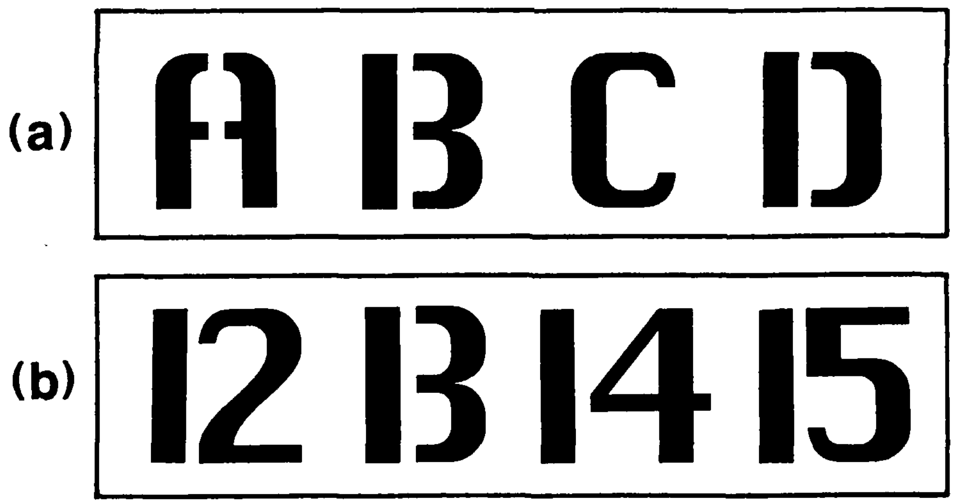


Figure 5.2 Example of use of contextual cues to identify an ambiguous signal (original figure)

now appears to be a 13. If you only saw Figure 5.2(b), you wouldn't think that any of those numbers were ambiguous, yet the "13" and the "B" are exactly the same. This is an example of the use of contextual cues to identify an ambiguous signal. When surrounded by the letters "A", "C", and "D", we see a "B"; when

surrounded by the numbers "12", "14", and "15", we see a "13". There are many studies that show that an appropriate context aids our ability to identify visual stimuli. For example, lines are easier to identify when they are presented in the context of an object, such as a box, than when they are presented alone (e.g., Weisstein and Harris, 1974). Letters are easier to identify when they are presented in a word than when they are presented alone (Reicher, 1969).

Palmer (1975) showed subjects pictures of a loaf of bread, a mailbox, and a drum. The bread and the mailbox were physically very similar. The subject's task was to decide which of the three pictures they saw. The subjects saw the pictures for such a short period of time that they could not be sure of which picture they saw. Sometimes, before seeing one of these pictures, subjects were presented with a scene such as a kitchen scene (i.e., a picture of a kitchen counter with utensils, food, etc.). When subjects saw a scene that was appropriate for the target picture (such as seeing the kitchen scene before seeing the loaf of bread), accuracy was significantly better than where they saw nothing before seeing the target. Performance suffered when subjects were "led down the garden path" with an inappropriate context and a target object that was physically similar to an appropriate object. For example, after seeing the kitchen scene, many subjects were sure they had seen the loaf of bread even if, in fact, they had been shown the mailbox.

In most cases, context helps or hurts us by setting the stage for expectations. When what we see or hear is compatible with what we expect, we process the information quickly and accurately. When it is incompatible, performance suffers. Examples of this can be found in videotapes of simulation studies where pilots say what they are thinking throughout the session. In an early TCAS simulation study, one pilot saw the traffic display and was so convinced that a "climb" advisory would follow that he never heard the many repetitions of the "descend" command (See pp. 313-314 for a detailed discussion.)

Our pattern recognition system is set into motion every time our senses perceive something. It is the first step toward processing complex information and problem solving. It is important to understand that pattern recognition cannot be considered in isolation. When we want to know how easy it will be to see or hear a particular stimulus (whether a simple line or tone or a complex message), we must consider the physical attributes of the stimulus, the context in which it will be presented, and the knowledge or expectancies of the perceiver.

Speech Perception

One example of complex pattern recognition is the comprehension of speech. Speech perception is a very interesting problem. Almost any small computer is capable of producing intelligible speech with the appropriate software and hardware. Nevertheless, it is incredibly difficult to get even the most sophisticated super computer to understand what the small stupid one said. These computers fail almost completely when they listen to a variety of human speakers say a variety of different things.

The French equivalent of Bell Laboratories has developed an automatic telephone where the caller speaks the number into the phone rather than dialing it. It works remarkably well with one notable exception. The phone does not usually work for Americans or other non-native French speakers, even though they may speak French very well. It appears totally unable to process the call. Why can't this computer recognize American French as well as French people can? The speech recognition systems that work best are "trained" to individual speakers who use a limited vocabulary. The speaker says the words to be used into the computer several times. The computer system then "learns" to recognize this limited set of words under ideal conditions. One necessary condition is a quiet environment since the computer can't differentiate between speech sounds and similar noises. Once the speech recognition system is trained to a speaker, it cannot tolerate much change in the speaker's voice, such as the rise in pitch that is often induced by stress.

To understand why speech recognition is so difficult, we must first examine the complexities of the speech signal. A spectrogram is a physical representation of the speech signal. It plots the frequencies (in Hz) of the speech sounds as a function of time. An examination of a spectrogram of normal speech reveals that it is impossible to say where syllables begin and end; words can only be differentiated when they are separated by silent pauses and these pauses do not always exist in natural speech which is quite rapid. This presents a problem for computers, since they are limited to the physical information in processing speech. We, on the other hand, use our knowledge of language to help parse the acoustic signal into comprehendible units such as words.

Another problem for speech recognition systems is the tremendous amount of variability in the speech signal. Ask one person to say "ba" five times. And these five simple sounds will all be slightly different (e.g., in terms of how long before the vocal folds vibrate after the initial release of the sound -- the initial opening of the vocal tract at the region of the lips. When these sounds are produced in context, they are even more variable. The "ba" in "back," for example, is slightly different than the "ba" in "bag."

There is even more variability from speaker to speaker. An examination of a physical representation of different English vowel sounds spoken by several native English speakers reveals a tremendous amount of overlap (Peterson and Barney, 1952). In many cases, it is only context that allows us to differentiate one from the other. This type of variability increases further if we include non-native English speakers. Being a non-native speaker affects not only how we produce speech sounds but also how we hear them. Unless we are exposed to the subtleties of the speech sounds as youngsters, we do not develop the capability to use the cues to the differences between these sounds in a speech context. The most famous example of this is the ra/la distinction. This distinction is used in German and English, for example, but not in many Eastern languages including Japanese. To native Japanese speakers, who learned English from other native Japanese speakers, "ra" is the same as "la" and "la" is "ra." They cannot distinguish one from the other even though they can distinguish the acoustic cues that differentiate these sounds for native English speakers when they are presented outside of a speech context (Miyawaki et al, 1975).

There are several other factors that influence our reception of speech sounds. One obvious one is the signal-to-noise ratio. In a noisy environment, some of the critical speech information can be masked. Generally, as the noise level increases, intelligibility decreases markedly. Specifically, the sounds that will be masked are the sounds of the same or nearby frequencies that exist in the ambient noise. Two other factors that have an additive effect on the effect of noise are the rate of speech and the age of the listener. When a person speaks quickly in a noisy environment, much more information is lost than when a person speaks quickly in a quiet environment or speaks slowly in a noisy environment. The effects of age on speech perception are two-fold. First, there is a loss of sensitivity, particularly to higher frequencies, that makes it more difficult to hear certain speech sounds. There is also a more subtle and intricate loss in sensitivity. After about age 50 we see a spreading in the widths of critical bands. This further compromises our ability to differentiate the speech signal from ambient noise. One result is that it is difficult to hear casual conversation at a noisy gathering. What do we do when we miss a word or part of a word? Based on context and our knowledge of language, we fill in the blanks - and we do so with utmost confidence. Studies have shown that if part of a word in a sentence is replaced with a noise, such as a cough or tone, the listeners fill in the missing syllables when they are asked to repeat what they heard. They are not able to locate the noise in time, even though they expect the noise somewhere in the sentence (Warren, 1970, Warren and Obusek, 1971).

To add to the problems of a 55-year-old (e.g., pilot) in a noisy environment (e.g., cockpit) trying to attend to a fast-talking speaker (e.g., controller), devices that transmit speech sounds, such as telephones, radios and headphones, selectively attenuate certain frequencies. The best earphones transmit everything from 25 to 15,000 Hz. These earphones wouldn't be very useful in the cockpit because the radios don't come close to this level of fidelity. Some of the frequencies that are lost (usually above 3000 Hz) are likely to contain some speech information since these frequencies are within the speech domain.

While many factors (e.g., age, noise, and transmitting devices) can degrade our ability to understand speech, there are very few factors that can destroy it. One thing that can, however, is delayed auditory feedback more commonly referred to as an echo. It is very disruptive for a speaker to have to listen to his or her own speech slightly delayed. Similarly, if we present speech in one ear and the same speech slightly delayed (beyond 30-40 msec) in the other, it makes the listener distressed and unable to understand the message. Delays below 30 msec. aren't as disruptive to comprehension, but are annoying and distracting. A study conducted with air traffic controllers showed that even a 5 msec. delay can be annoying (Nadler, Mengert, Sussman, Grossberg, Salomon, and Walker, unpublished manuscript). Fortunately, this is an artificial situation (i.e., induced by equipment) that can usually be avoided.

It is almost amazing that we understand speech as well as we do. The speech signal is incredibly complex and often embedded in noise. Yet, under most circumstances, the system works very well and failures to comprehend spoken messages are the exception rather than the rule. Unless the workload and stress levels are terribly high and/or the environment is excessively noisy, we usually do OK. Armed with our knowledge of language and aided by context, we are able to decipher the signal and understand the message. And then, sometimes, we just fill in the blanks.

Memory

Memory is a key component in our information processing system. Simple recognition requires that the pattern in front of us match a pattern in memory and most complex problem solving requires applying information stored in memory to the task at hand.

The Sensory Store

Scientists usually think of memory as three different memory structures: sensory, short-term (also called working), and long-term. Table 5.1 summarizes the key characteristics of these three structures. A sensory memory structure probably exists for each of the five senses. These five sensory modalities take in information automatically; there is no way to avoid it. If you open your eyes, information comes in. Unless you plug your ears, auditory information enters. This information that enters sensory memory automatically cannot be maintained intentionally. You can only look again or listen again to the same message. Otherwise, the scene or message is gone in a short time from five-tenths of a second to two seconds. For some auditory information, sensory memory has been demonstrated to be about a quarter of a second which is the length of most syllables. Our capacity for sensory storage is very large. The information is held in an unprocessed mode. The meaning of a word, for example, is not yet accessed. The information must proceed to short-term memory with the aid of pattern recognition procedures for further processing.

The sensory store takes in a lot of information but holds it for such a short time that only a small portion of this information can be recognized and transferred to short-term memory, and thus, available for further conscious processing. The rest of the information is lost and this loss usually goes

Table 5.1
Memory Structures

FEATURES	SENSORY	(WORKING) SHORT-TERM	LONG-TERM
Information input	Automatic	Requires Attention	Rehearsal; Higher order processing
Information duration	0.5 to 2 sec	20-30 sec	Decades
Information capacity	Large	7+ or -2 items	No known limit

unnoticed. Sperling (1960) conducted a series of experiments that demonstrates the capacity of this sensory store. In one experiment, he showed a card with a four-by-three matrix of letters and numbers (Figure 5.3) to subjects for 50 msec. When subjects were asked to recall all the letters and numbers they saw, they remembered seeing twelve but could only name three or four. Were more items available in sensory store but lost before they could be reported? To investigate this, Sternberg showed the same type of matrix for the same amount of time, but this time, he also played a high-, medium-, or low-pitched tone. If the tone was high-pitched, then the subject was to report the top row. If the tone was medium-pitched, then the subject was to report the middle row. If the tone was low-pitched, then the subject was to report the bottom row. The tone was played immediately after the display disappeared. The subjects were asked to report only the letters and numbers that had appeared in that row. In order for subjects to report all four items in the row correctly, the full array of twelve items would have to be available in sensory memory when the tone sounded. This was, in fact, the case. The subjects were able to recall all four letters and numbers, no matter which row was cued. Without the cue, however, most of the items were "lost" before they could be reported.

Sensory memory has also been demonstrated in the auditory domain. With the use of earphones, we can present letters or digits that appear to come from three different places. For example, in the right ear, we present "1, 2, 3" and simultaneously present, i.e., superimpose "4, 5, 6". In the left ear, we present "7, 8, 9" with the same "4, 5, 6" superimposed. What the subject "hears" is "1, 2, 3" in the right ear, "7, 8, 9" in the left ear, and "4, 5, 6" in the center of the head. If one of these locations is randomly cued after presentation, recall for the numbers presented there is nearly perfect. Without the cue, only three or four of the numbers can be recalled. In sensory memory, much visual and auditory information is stored but lost quickly. A small proportion of the stored information is transferred to short-term memory for further processing.

THE STIMULUS CARD

7	1	V	F
X	L	3	5
B	7	W	4

Figure 5.3. Example of a four-by-three matrix of letters and numbers shown to subjects to illustrate sensory store capacity of short-term memory. (original figure)

Short-Term Memory

Our second memory structure is working or short-term memory (STM). We can think of the information stored in short-term memory as what is immediately available in consciousness. It is what we are thinking about at the time. Maintenance of information in short-term memory requires attention. That is, if you want to keep information available here you must focus on it or use it in some way. How many times has someone introduced you to someone and one minute later you can't recall the name? You heard the name clearly, but you didn't perform any cognitive effort to process the information. Unlike the information in your sensory store, you can keep the information in short-term store by rehearsing it, that is, repeating it. Without rehearsal, the information will be available for only 20 to 30 seconds. Even with rehearsal, the information in short-term memory is fragile. If someone tells you a phone number, repeating it will keep it available on your way to the phone. If someone approaches you while you're rehearsing and asks you the time, your response of 3:45 could displace the phone number out of STM.

The information in STM is very susceptible to interference. The more similar the interfering information is to the information in STM, the stronger the interference will be. For example, numbers can displace other numbers more easily than names can displace numbers.

The capacity for storage in short-term memory is relatively small: seven items, plus or minus two (five to nine) items. The "items" can be digits in a phone number, for example, or they can be packages or chunks of information. If someone read a string of letters such as "F, C, J, M, U, B, I, F, T, H, F, V, K, A, I", then asked you to recall them, you would probably be able to recall between seven and ten of them. If the same letters were read in logical groupings, such as "FBI, CIA, JFK, MTV, UHF", you would probably be able to recall all of them. Fifteen items have been "chunked" or grouped into a meaningful set of five. Similarly, "2, 0, 6, 3, 8, 4, 5, 7, 9, 1" will be easier to recall as "206, 384, 5791", particularly if it is a familiar phone number. These two examples illustrate two points. First, the capacity of short-term memory is increased when the information is organized. Second, if the information to be stored in STM is familiar, that is, already exists in long-term memory, then it will be easier to maintain in short-term memory. It should be noted, however, that the definition of a "chunk" of information can be arbitrary. For example, whether a radio frequency can be considered as one chunk of information or as four separate pieces of information is debatable (and should probably depend on whether or not the frequency is a familiar one).

Long-Term Memory

Long-term memory is the "warehouse" of information stored up over a lifetime. There is no known limit to the amount of information we are able to store in long-term memory or to the length of time we are able to store this information. Information is rarely lost from memory, but it is frequently more difficult to retrieve than we would like. Often, we know we are very close before we successfully access the information. For example, in trying to recall a name, we may be able to recall what letter it begins with, the number of syllables, or what the name "sounds like" but the actual name escapes us. This is called the "tip-of-the-tongue" phenomenon. The information is in long-term memory but we can't recall it into short-term memory at that moment. Eventually, we are usually able to reconstruct the name from the descriptive information that we can retrieve.

Much of memory is ***reconstructive***. Information may be available even though it isn't encoded in the same form as the information for which you're searching; it may have to be derived. For example, the number of rooms in the house that you lived in when you were five years old is not something you consciously stored. It is also not something that you can recall quickly. However, you probably are able to recall an image of the house and "walk through" and count the rooms.

Memory is also constructive in the sense that we not only store information that is given directly to us, but we also store whatever that can be derived from that information. Bransford, Barclay, and Franks (1972) read many sentences to subjects in their experiments and later asked them if the test sentences were ones they had heard before. They found that subjects could not distinguish between the sentences they heard and ones that could be logically inferred from the ones they heard. It was the processed meaning of the sentences, not the specific words, that was stored in long-term memory.

There is some physiological evidence for the existence of short- and long-term memories as separate distinct structures in the brain. The following case illustrates this point. H.M. incurred brain damage as the result of an accident. Because of this damage to the temporal lobes, H.M. was unable to transfer information from short-term into long-term memory. The information stored in long-term memory before the accident remained intact and could easily be recalled. This, along with his functioning short-term memory enabled H.M. to carry on normal conversations with his doctor and others. Without the ability to transfer the information to long-term memory, however, the conversations were forgotten. If the doctor left the room and returned minutes later, there was no evidence that H.M. had any memory of the conversation that took place just minutes before.

With disease such as Alzheimer's, there is also evidence of a separation of short- and long-term memory. In the beginning stage of the disease, transferring information from short-term memory into long-term memory is problematic. Later, long-term memory degenerates and eventually the disease invades so deep into the memory system that even language can be forgotten.

In the absence of brain disease or damage, there are things we can do to help store information effectively in long-term memory. If the material to be learned can be organized around existing knowledge structures, (i.e., information already known), then it will be more efficiently stored and, thus, easier to recall. It is easier to learn more about something you already know than to learn the same amount of material about something totally foreign or to learn it as isolated facts. Cognitive effort can also help to store information in long-term memory. This effort can be intentional or incidental. We can study to memorize facts (intentional) or we can use information so often, e.g., a phone number, that we learn it whether or not we intend to do so. On the other hand, information that we would like to keep easily accessible (such as memory items on a checklist) may not be readily available without regular review. Our memory for important,

complex information that is not used regularly but does need to be quickly accessed, requires periodic maintenance - particularly if this information is expected to be recalled in stressful situations.

Chapter 6

Display Compatibility and Attention

by Christopher D. Wickens, Ph.D., University of Illinois

Display Compatibility

As we follow the sequence through which information is processed by the pilot, the first critical stage is that of ***perception***, that is, interpreting or understanding displayed information. However, there are features in display design that can allow this interpretation to proceed automatically and correctly or, alternatively, to require more effort with the possibility of error. This is the issue of the compatibility between displayed information (stimulus) and its cognitive interpretation. Based on that understanding, a response is triggered. Compatibility generally refers to the relationship between a display's representation, the way in which the display's meaning is interpreted, and the

way in which the response is carried out. ***S-C compatibility*** refers to the relationship between how a stimulus changes on a display, and how it is to be cognitively interpreted. ***S-R compatibility*** refers to the relation between displayed stimulus change and the appropriate response. The important design issue in S-C compatibility, which we shall now consider, is whether the change in a display state naturally fosters the correct cognitive interpretation. We provide several examples below.

Color is one important component of display compatibility. When a display changes color, does the color on that display immediately give the correct interpretation to the pilot of what that color is supposed to mean? The meaning of certain colors is related to ***population stereotypes*** which must be kept in mind by designers. A designer might think "I have a meaning I want to convey, what color should I use to convey that meaning?" This is really working backwards because it does not address other population stereotypes a color might have. What the designer really wants to do is say: "When a color appears on a display, what will the pilot automatically interpret it to mean?" The problem occurs when colors have multiple stereotypes, and so the pilot may instinctively interpret one that is different from what the designer intended. Red has a stereotype of both "danger" and "stop" or "retard speed." Now a pilot sees red, in the context of airspeed control. Does it mean "slow down" or does it mean that "airspeed is already too slow and there is danger of a stall?" Possible conflicts of color stereotypes must be carefully thought through by the designer, to make sure that a given color has an association that can't possibly confuse or be confused and trigger the incorrect interpretation.

The second component of display compatibility is the spatial interpretation of display orientation and movement. This relates to the movement of a display and how a pilot interprets what that movement signals. Roscoe (1968) cited two principles that define display compatibility. The first is the ***principle of pictorial realism***. The spatial layout of a display, that is, the picture of a display, should be an analogical representation of the information it is supposed to represent. The second principle that helps define display compatibility is the ***principle of the moving part***. The moving element of a display should move in the same orientation and direction as the pilot's mental model of systems moving in the real world.

A good way to illustrate these two principles is with examples of hypothetical airspeed indicator designs as shown in Figure 6.1. These are not necessarily the ideal ways of designing an airspeed indicator, but they either confirm or violate the principle of pictorial realism or the principle of the moving part. A pilot's mental model of airspeed is something with a "high" and "low" value. Therefore, according to the principle of pictorial realism, a vertical

representation is more compatible than the circular one as shown in design (d).

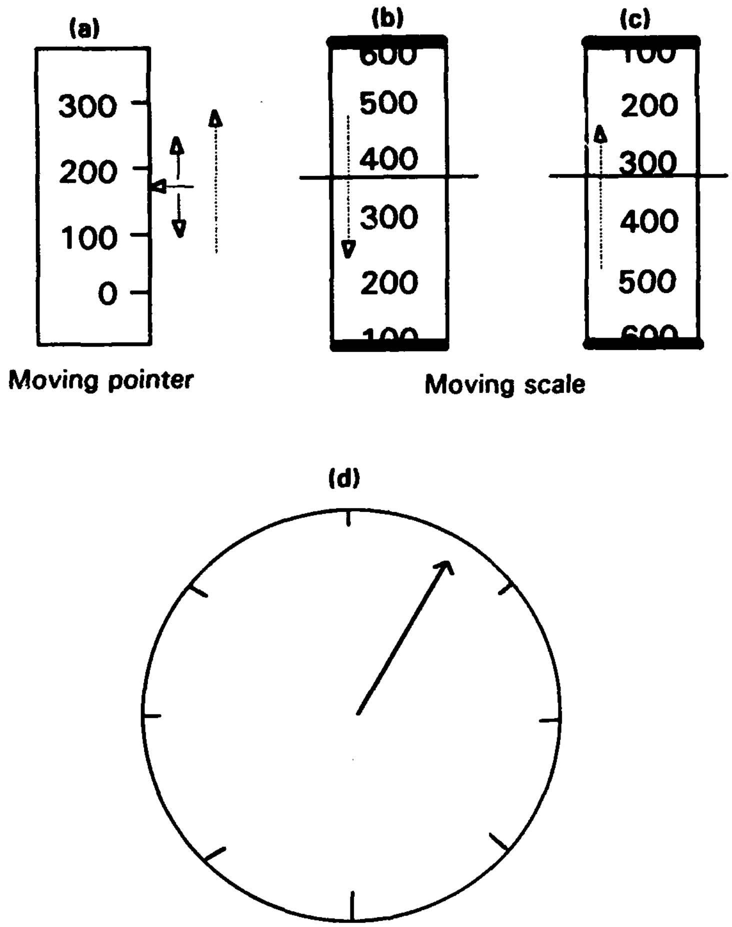


Figure 6.1. Different altimeter displays illustrating the principles of pictorial realism and of the moving part. (from Wickens, 1992)

Also, our mental model has high airspeed at the top and low airspeed at the bottom. So a fixed scale moving pointer indicator with the high airspeed represented at the top, as shown in design (a), is compatible with the principle

of pictorial realism, whereas a moving scale indicator with the high airspeed represented at the bottom violates that principle (d).

Consider the display of altitude as another example. There has been a good deal of research suggesting that pilots think of the aircraft as the moving element through the stable airspace, not as the stable element in a moving airspace (Johnson and Roscoe, 1969). So when an aircraft gains altitude, it is compatible with both the principle of the moving part and the principle of pictorial realism for the moving part of the display to move upwards, and when the plane descends, for the moving part to move downwards. This is exactly what we get with a fixed scale moving pointer display (a). You have high altitudes at the top, low altitudes at the bottom, and your moving pointer is in a direction of motion that is compatible with the pilot's mental representation of what is happening in the environment. That is, it conforms to the principle of the moving part. With a fixed pointer moving tape display, there are two possible design orientations. The situation in design (b) has the high altitude at the top of the tape and the low altitude at the bottom; again, conforming to the principle of pictorial realism. But an increase in altitude is signaled by a downward movement on the display--a violation of the principle of the moving part. The alternative is to present the low altitude at the bottom of the tape and the high altitude at the top. In that case, when the plane climbs, the tape moves upwards, and you've satisfied the principle of the moving part but violated the principle of pictorial realism. This is one of those cases of competing principles.

While it would seem therefore that the fixed scale display is ideal because it conforms to both Roscoe's principles, it turns out that even this is not necessarily the ideal solution because of a problem with scale resolution. For variables like altitude, you can't print the whole scale unless it is printed so small it is nearly impossible to read. That is the nice thing about moving scale displays. They can accommodate a much longer scale because they are not constrained by space. A compromise solution which could be adopted here is called "frequency separation," in which the pointer moves rapidly across a fixed, partially exposed scale to reflect high frequency changes. But lower frequency, longer duration changes that will require exposing a different scale range are accomplished by moving the scale.

Attention

Attention may be characterized as a limited capacity available to process a lot of information. Our discussion of attention here will lead in two directions: discussing the principles of multi-element display design, and the use of head-up displays. Then in Chapter 11, we shall discuss the issue of dividing

attention when trying to perform several tasks at once, and measuring the attention demands of tasks: the issue of pilot workload. The issue of attention can really be divided into three different aspects of human abilities. One aspect is ***focused attention***--how easily we can focus on one source of information and ignore the distraction of other information. Successful focus is the opposite of distraction. Another aspect of attention is ***divided attention***--how easily we can divide attention between two activities and do two things at once, or process two display channels at once. These activities could involve the pilot flying at the same time he or she is communicating, or perceiving vertical velocity at the same time that heading is perceived. Finally, we have the aspect called ***selective attention***, and this describes how easily and how carefully the pilot selects particular channels of information to be processed at the right time (e.g., is the pilot sampling an instrument when he should be looking outside, or attending to data entry on the FMC when he should be attending to airspeed control).

Focused Attention

A discussion of focused attention and distraction leads to consideration of the electronic display issue. One of the things that we know from basic psychology is that all information that falls in about one degree of visual angle is going to get processed whether you want it processed or not. We know in aviation displays that clutter is going to be an *inevitable* consequence of putting more information in a smaller and smaller space. This will be important in the discussion of head-up displays to follow. The issue now is to minimize the confusion caused by clutter, and images that are too close together in the visual field. How can we increase the pilot's ability to focus attention on one displayed item and ignore other things that may not be relevant? We are finding in research that **color is an extremely useful technique for segregating different sources of information.** Coloring all of one type of information in one color and different information in a different color can allow us to focus in on, say, all of the information that is of one t   , and ignore the information that is of the other, even if they are in the sam spatial location.

With auditory messages too, the issue of confusion and distraction is relevant. How do we allow the pilot to focus attention on one auditory channel of information (say a synthesized voice message from the cockpit), while filtering out conversation from the copilot or controller, so that the latter will not get confused with the cockpit alert? The answer here is again in terms of physical differences, in this case making messages sound as different from each other as possible -- perhaps by purposefully making computer-driven messages sound artificial.

Divided Attention

When we consider divided attention, particularly attention divided between different aspects of a display, designers are interested in creating for the pilot a sense that two (or more) parts of a display that are to be related can be perceived at the same time. This objective can sometimes be achieved by bringing them close together in space. This, of course, is the principle underlying the development of the head-up display. Also, any sort of static display ought to have the labels of an indicator very close to the indicator's actual moving part. In fact, the analysis of the USS Vincennes incident when the Navy ship shot down the Iranian airliner, revealed that the label on the Navy's radar system that indicated whether the altitude was increasing or decreasing was considerably separated from the actual indicator of XY position itself. So the separation of these two pieces of information may, in part, have caused the controllers on the radar display to misinterpret what that altitude trend information was showing, assuming that it represented a descending attacking fighter, rather than a climbing neutral airliner.

Of course, spatial closeness can be overdone. As we noted above, too much closeness can create display clutter and thereby be counterproductive. Thus, relative closeness between related display channels is probably more important than absolute closeness.

In addition to spatial closeness, it is also possible to **use a common color to bring together in the mind two things that may be spatially separated**, and make it easier to divide attention between them. As we note in the next chapter, for example, it may be useful to use a common color to show the relationship between a display and its associated control, when these are not colocated; or, in an air traffic status display, to code all aircraft with similar characteristics (e.g., common altitude) with the same color. Because color can be processed in parallel with other features of a display, it is often useful to use the color coding of an object to facilitate divided attention.

A third display feature that can improve the ability to divide attention between two indicators is to present them as two dimensions of a single object. Perhaps the best example of this is the attitude display indicator (ADI) that represents two independent dimensions of flight control. It represents both pitch and roll as the vertical location and the angle of the horizon. That design feature greatly improves the ability to divide attention between those two critical pieces of flight information for integrated lateral and vertical flight control.

Another important way of designing displays to facilitate parallel processing is through the creation of ***emergent features***. These are perceptual characteristics of

a set of displays that are not the property of any single display. A good example of an emergent feature is the imagined horizontal line, that connects the tops of four vertical column engine indicators on a four-engine aircraft, when all engines are running at the same level as shown in Figure 6.2.

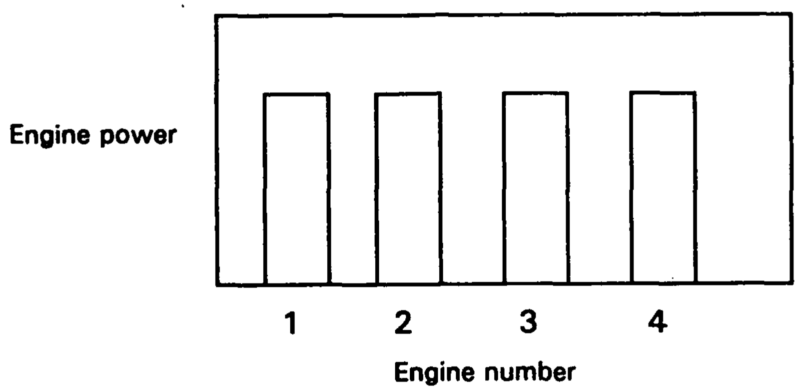


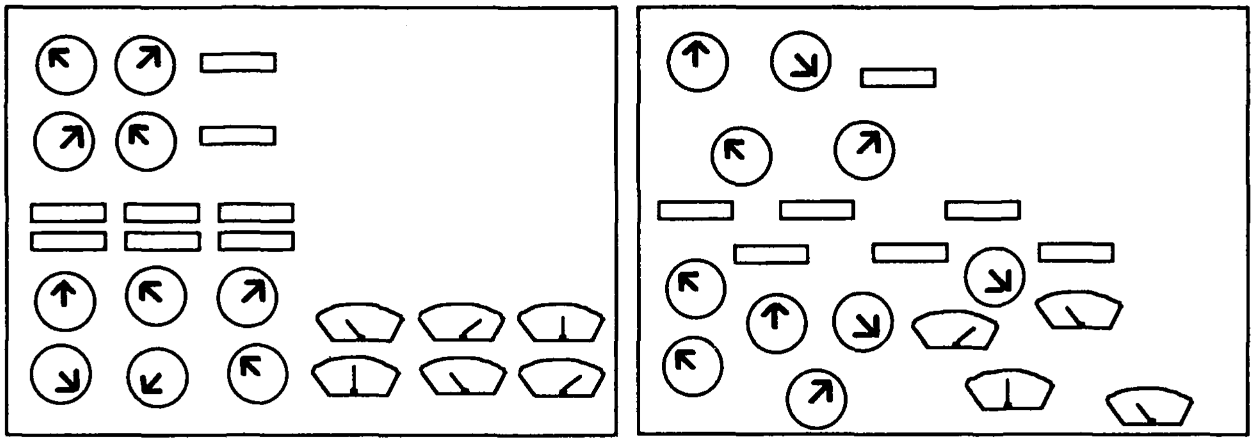
Figure 6.2 Vertical column engine indicators for a four-engine aircraft. (Wickens, 1992)

Two other characteristics that will improve the ability to process information in parallel will be discussed in more detail in our later section on workload. These are the ***automaticity*** with which information is perceived (the more automatically we process one symbol or piece of information, the better we can do so in parallel with other display processing), and the use of ***separate modalities*** of information display (i.e., auditory and visual channels).

Selective Attention

The pilot's ability to select information that is needed on the display at the appropriate time can be improved by three factors. First, and most obviously, ***training*** can improve selective attention. There is reasonably good evidence that pilot's ***scan patterns*** (good indices of what is being attended when), change as a function of their skill level, indicating an evolution of selective attention ability. Second, ***display organization*** provides a good way of enabling the pilot to find (look at) the information needed at the right time. One can contrast the more organized display in Figure 6.3a, with the less organized one in Figure 6.3b, to see the difference. **However, it is important that the physical organization of the display be compatible with the mental organization that defines the pilot's**

information needs. That is, displays that are clustered or grouped together would be those that are also used together.



(a) (b)

Figure 6.3 (a) Example of good display organization. (b) Example of poor display organization. (Wickens, 1992)

Display consistency is a third variable that effects the pilot's ability to selectively attend to the right sources of information at the right time. Where possible, similar types of displays should be located at similar places, across different viewing opportunities. This applies both for display locations across different cockpits, and for multifunction displays across different pages that may contain similar material. Finally, as we described above, ***display clutter*** will be a hindrance to effective selective attention. It is difficult to visually find what you want on a cluttered display.

Head-Up Displays

The design and use issues of the head-up display highlight many of the issues of attention discussed in the previous pages. Figure 6.4 shows a sample of a head-up display (HUD) developed by Flight Dynamics, Incorporated. It is flown in Alaskan Airlines planes. The HUD was designed primarily to bring visual channels closer together in space so as to improve the ability to divide attention between them. Instead of having critical flight instrumentation physically separate from the outside world, the HUD overlays certain aspects of this information on the view of the outside world. The goal of the head-up display

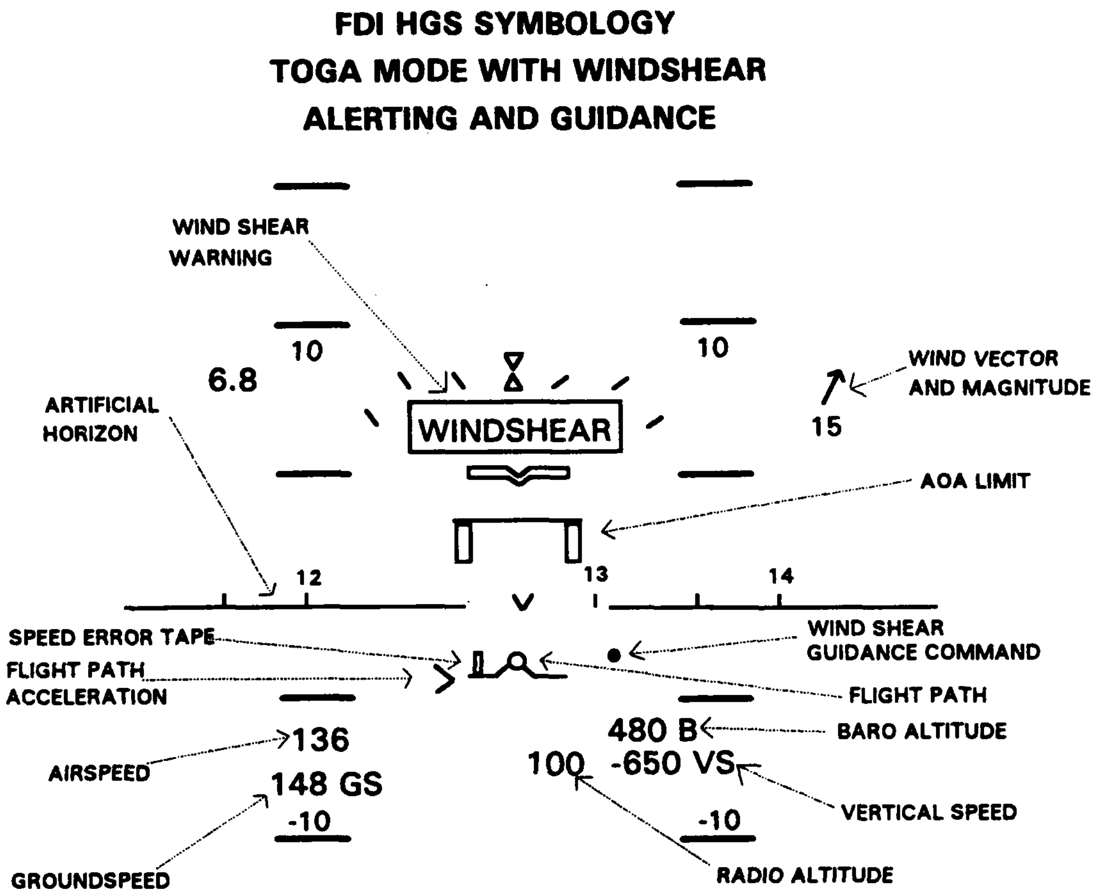


Figure 6.4 Sample of head-up display (HUD). (Desmond, 1986)

is twofold. One, as noted above, is to reduce the need for visual scanning between instruments and the outside world. The second goal is to portray certain critical pieces of information that conform with the environment so they can be directly superimposed on that environment. These would include, certainly, the runway symbol, the horizon line, a flight path representation, and a symbol of the aircraft's current and predicted position. This ***conformal symbology*** then can be interpreted by the pilot as belonging at locations along his or her line of sight beyond the HUD.

HUD display development and research has a very long history in the military. There are a number of issues in the military, like flying inverted and getting out of high-G combat situations, that are less relevant for the design of civil aircraft. On the other hand, it has been recently introduced and successfully

flown in Alaskan Airlines planes and has had very good reception (Steenblik, 1990). The first category three landing was at Seattle-Tacoma Airport in late 1989. The pilots who have flown with it generally have liked it and have found that it does a good job of allowing maneuvering and landing in very low visibility conditions. At the same time, it keeps them actively involved in the control loop rather than turning over control to automatic landing systems, thereby maintaining a level of involvement which pilots generally value. Flight tests with the HUD have been quite successful. Figure 6.5 shows an example of the "footprints" of landing touchdowns made on a series of category one and category two landings done in simulations with and without a HUD. It shows greater touchdown dispersion without the HUD than with it. It also tells us that there were six go-arounds in the approach without the HUD and no go-arounds with the HUD. Desmond (1986) reviewed the development of the HUD and its implementation in the aircraft.

The critical issues in HUD design relate not so much to whether they are a good thing or bad thing, although some researchers have phrased it that way, but rather to the appropriate design guidelines to follow, how HUDs can be improved, and to identification of the potential pitfalls in HUD use (Weintraub & Ensing, 1992).

In the analysis of HUDs, there are three conceptually different domains. One domain has to do with the optics of the HUD, that is, how they are ***collimated***, how the lenses are configured, and where they are located (the visual angle between the HUD instrumentation and the line of sight out the cockpit toward the runway during approach). A second is the ***symbology*** of the HUD. What exactly should be placed on the HUD, and in what format? How much of this should be nonconformal symbology? The third domain concerns the whole issue of ***pilot attention*** in the HUD. How does human attention switch back and forth between the HUD instrumentation and distant objects in the far environment? How well can human attention be divided between instrumentation and things in the far domain? What are the consequences of focusing attention on the near HUD and ignoring information that is out there in the environment?
In addition to these three issues of HUD research, there are four important categories of differences between typical HUDs and conventional flight instruments. First, HUDs are, of course, displaced upwards to overlap the visual scene. Second, conventional displays are presented at a short optical distance. HUDs are typically collimated out to near optical infinity. Third, there are significant differences in the symbology between conventional instruments, which often, although not necessarily, have an older round dial symbology, and HUD instrumentations which typically have a much more novel symbology. Fourth, the different symbologies represent the movement of the airplane differently. Most conventional instrumentation for presenting guidance

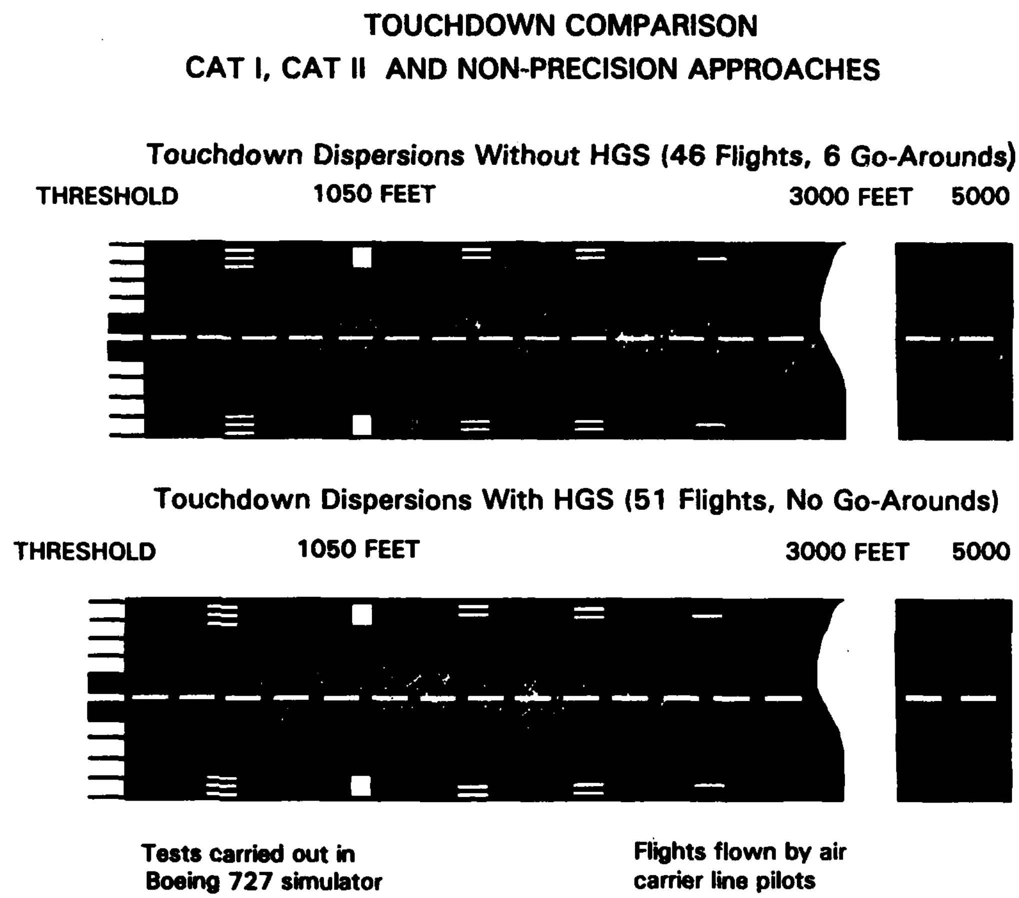


Figure 6.5 **Touchdown dispersions with and without HUD for nonprecision approaches. (from Desmond, 1986)**

information is based on the relationship of the airplane to the air mass. Some HUD symbology (e.g., that used by Flight Dynamics), in contrast, may be based on the inertial guidance of the plane and therefore provides information with respect to the ground surface. Differences in flight test performance between HUD and conventional instrumentation could result from any or all of these differences in design features.

HUD Optics

When we view objects up close, the light rays from the object hit the eyeball in a converging orientation. They are not parallel. The muscles surrounding the lens must activate or "refract" to bring that image into focus. For objects more than five or six meters away, the light rays travel in a roughly parallel orientation. The lens relaxes its shape and the more distant object is brought

into focus. This change in lens shape we call ***accommodation***. Accommodation is not instantaneous, which is why we have a difficult time going from viewing something far away to suddenly reading something up close. This problem with accommodation increases with age. The goal in the design of the HUD is to present the information, which is superimposed on the windscreen, so the light rays travel in parallel to the eyeballs, and so the information is essentially perceived as being out at a great distance (i.e., at optical infinity). This is all done by a series of collimated lenses down at the bottom of the HUD that take the image generated on a CRT and transform it into parallel rays. Hence, the rays from the far domain of the runway or distant aircraft and the rays from the near domain from the instrumentation are all displayed in parallel. Information from both domains therefore requires very little accommodation at all.

In making any sort of comparison between the HUD and conventional instrumentation, one of the issues is the fact that conventional instrumentation is usually presented at close range while head-up display information is presented out at optical infinity. There has been some dispute in the human factor literature regarding whether or not it is appropriate to collimate the HUD instrumentation out towards optical infinity. The issue is conceptually simple. At different times, the pilot has to have the eyeball accommodated to two different distances. On the one hand, he has to look out of the cockpit and focus on the things that require far accommodation like the runway, distant aircraft, targets in space, etc. On the other hand, the pilot has to spend time looking at close things, particularly airport approach plates and maps in the cockpit. So, a decision must be made about where to put other aspects of the critical flight information. Should it be projected in close, where processing is more compatible with the maps, or projected "out there," where processing is more compatible with the distant world? The general guideline followed by HUD designers seems to be that it is more important for the pilot's eye to be well accommodated to the distant features. Therefore imagery is either collimated out to optical infinity or a little less than that, but still fairly far out, which keeps the light rays almost parallel.

Despite the decision which has been made for pilots to view HUD instrumentation at optical infinity, there isn't a lot of data to suggest how pilots really do accommodate back and forth between the "near" and "far" domains. One of the few studies in this area to date has been done by Weintraub, Haines, and Randall (1985). They used a static test in which they examined the pilot's ability to switch between near information and far information. The near information was a digital altitude and air speed display on a HUD. The far information was the presence of an X at the end of a runway, which would signal that the runway was closed (Figure 6.6). The experiment would present

the HUD information, then would suddenly present the runway information, and determine how long it took the pilot to confirm appropriate altitude and airspeed, and then make the decision about whether the runway was open or closed. Essentially they were asking the pilot to switch attention from the near domain, (the air speed and altitude), to the far domain, and then make a response of whether there was an X present or not. In one condition of their experiment, the instrumentation was presented head down, and optically close. Therefore the pilots not only had to switch attention from the near to the far, but they had to accommodate from the HUD to the distant runway.

Figure 6.6 shows the results from this condition. The solid line represents the state of accommodation, changing from the near to the far symbology. This is

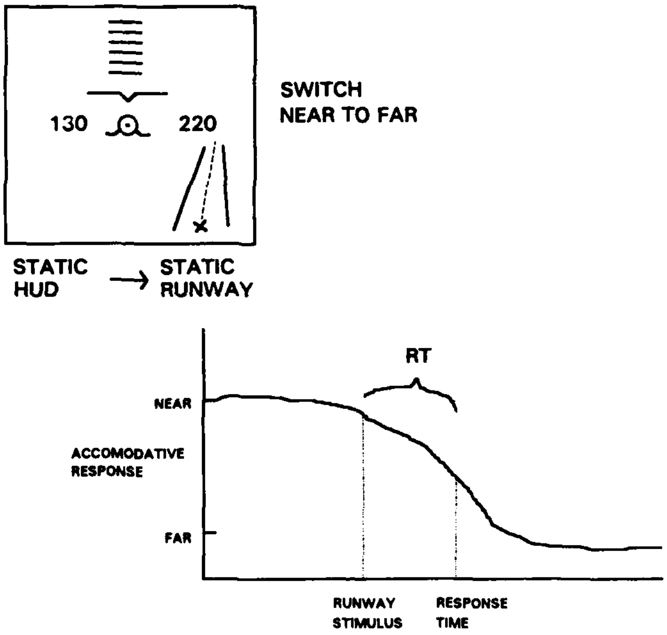


Figure 6.6. Example of HUD stimuli used in experiment, and graph showing results of tests of pilot's ability to switch from near to far information. (adapted from Weintraub, Haines & Randall, 1984)

called the ***accommodative response***. The important point to note in this figure is that the time to make this decision is influenced partially by how far they have to accommodate, but also they can make the response well before they have completely reaccommodated to the greater distance. This finding suggests that you don't need to have perfect visual information in the far domain before you are able to process it and use it. Nevertheless, this was the first experiment that

really documented a major cost to reaccommodate, and that cost showed up in performance. It is an experiment that strongly suggests the importance of keeping that imagery close to optical infinity rather than close in. Weintraub, Haines and Randall also varied the visual angle between the HUD and the runway information. They compared two conditions. In both conditions, the HUD imagery was collimated to optical infinity. In one condition, the HUD imagery was overlapping the runway and "head-up." In the other condition, it was not overlapping the runway and "head-down." In the head-down condition, the imagery was still optically far, but was no longer superimposed on the runway. Instead, it was positioned at the same location as the true conventional instrumentation. So to get information from the runway and HUD in the head-down condition, the pilot still had to visually scan up and down, but didn't have to reaccommodate. The investigators found almost no difference in performance between the head-up and head-down conditions in terms of the ability with which judgments could be made. These results suggest that the advantages in the head-up display may be more in the symbology on the one hand, and in lessening the need to reaccommodate, than in the fact that there is overlapping imagery.

Physical Characteristics

In addition to the physical and optical placement issues, there are a set of other physical characteristics of the HUD that are worth noting. Many of these are taken from a series of guidelines presented by Richard Newman, who did a fairly extensive review for the Air Force, and whose findings are applicable to civil aviation as well (Newman, 1985). One of the guidelines concerns the ***eye reference point***. It turns out that in viewing a HUD, the imagery changes and the ability to interpret it changes a little bit, depending on where the eye is positioned relative to the HUD. Newman argues very strongly that the HUD positioning should be adjustable to allow different seating postures, so it could be moved when the pilot is scrunched forward or sitting back. A second issue concerns the ***field of view***. That is, how much of the outside world should the HUD incorporate? A lot of technological effort has been put into designing HUDs that can present a wide field of view. One of the guidelines is that the field of view should be at least wide enough so that when you are landing into a crosswind with a very substantial crab angle, the runway is still visible on the HUD, even as the aircraft is crabbed maximally into the wind. This difference between aircraft heading and velocity vector indicates how wide the field of view should be on the HUD.

Another issue that isn't well-resolved concerns what happens when conformal symbology on the HUD moves out of the field of view. Suppose a pilot is flying directly towards the runway, and then changes course so that now the runway

symbol slides off to the side of the HUD. Should it disappear or just freeze on the side so the pilot still clearly perceives that it is off to the left or the right, but now perceives an underestimation of the magnitude of the deviation.

Another physical characteristic concerns the frequency with which the HUD is updated. For analog information on the HUD, a guideline is that the variables should be refreshed at around 10 to 12 hertz, sufficient to give good performance. For digital information on the other hand, you certainly don't want that fast updating, because digits tend to be unreadable. Therefore, something like 3 to 4 hertz is probably appropriate.

Symbology

The symbology issue can be broken into two major domains. The first relates to some of the sensory factors that relate to issues in visual and auditory perception. For example, what should be the intensity of the HUD imagery? How bright should it be? What is the necessary intensity to perceive across the conditions ranging from night viewing, in which you can get by with fairly low intensity, to incredibly bright snow cover? Is a single fixed intensity adequate, or should there be automatic or manual intensity control? A related issue concerns the ***transmittance***. Newman has recommended that no less than 70 percent of the outside world light should be transmitted through the HUD. Weintraub argues instead that it should really be more like 90 percent (Weintraub & Ensing, 1992). In fact, the Flight Dynamic HUD used by Alaskan Airlines has about 90 percent transmittance.

Color is another issue in HUD design. The current HUD designs tend to be monochrome (green). One of the reasons is that the monochrome display transmits a lot more light than a color HUD. Color of course has benefits, but color, as viewed on the HUD, may have some real problems in terms of interpretation, particularly when several colors are to be used. Under the varied conditions of illumination in which a HUD may be used, any more than four or five colors will create a real risk of confusion.

Cognitive issues in the design of HUD symbology are also relevant. The Air Force has done some good research in terms of the nature of the HUD symbology and how that can be best interpreted (Weinstein, 1990). The nature of the pitch ladder is one example. How do you make the pitch ladder as unambiguous as possible in depicting whether the aircraft is nose-up or nose-down? Here is where color comes in. One of the problems with the HUD is that its graphic representation of what is up and down is not as good as the colored representation on the typical Attitude Display Indicator using blue and brown.

There may be a role for color in HUDs to help make the simple discrimination of what is above and what is below the horizon.

The use of the inertia guidance system is an important cognitive issue. Its importance is suggested by the fact that the evaluation of the HUD flown in Alaskan Airlines revealed that the characteristic that pilots seemed to like most is the fact the guidance given by the HUD is based on inertial guidance rather than air mass guidance. In other words, the pilot actually gets a representation on the HUD instrumentation of where the plane is heading relative to the ground, rather than relative to the air mass through which it is flying. So this indicates that possibly the major benefits may be in what the HUD presents rather than where it is physically presented.

Some issues have to do with the development of flight director displays on HUDs. These correlate very closely with the same issues of the flight director for presenting head-down information. What is the appropriate tuning? What are the appropriate rules to guide the flight director?

One major symbology issue concerns how much information should be on the HUD. Should a HUD only present the necessary conformal flight information, the things that are necessary for actual flight path guidance, and, therefore, conform to (and can be superimposed on) the world outside? Should the HUD also present different kinds of flight parameter and alerting information, and if so, how much? As we see below, this impacts the issue of display clutter.

Finally, there is the issue of multimode operations. Some HUD designs present a lot of information in a relatively small space. If this is viewed as a problem, then designers often recommend that the pilots be given the option of calling up alternative forms of information. However, any time the designer creates multimode situations, you start dealing with problems of menu selection, forcing the pilot into computer keyboard operation. Such operations have a number of potential dangers at critical high workload times during the flight, when the HUDs are likely to be in use.

Attention Issues

The initial goal of the HUD was to resolve the problems of divided attention by superimposing the two images. Once that decision was made, then there followed the issue of how to improve the symbology, and the decision to collimate the images at optical infinity. The real question is whether or not simply superimposing images of nonconformal symbology does address the problems of divided attention, or whether it creates the potential for other problems.

There are three possible attention problems that are created by superimposing visual images. One is whether or not the resulting ***clutter*** disrupts the ability to focus attention. Are there problems trying to focus attention on the far world, (the runway out there) when there is a large amount of symbology in the near domain that may be partially obscuring it? Can these problems be addressed by reducing HUD intensity? The second problem, a related one, is related to ***divided attention and confusion***. If a pilot is actually trying to process the far-world information and the near-world symbology simultaneously, is there a possibility of confusion? For example, when the aircraft moves and the far-world runway then moves relative to the HUD, could the motion of the runway be misinterpreted as being part of the movement of analog symbology on the HUD? The third problem, related to ***attentional tunneling or fixation*** we now discuss in some detail, in the context of research at NASA Ames.

One of the few studies that has been conducted with a dynamic head-up display to examine attentional issues has received a fair amount of publicity, although it has some methodological problems. It is a study done by Fischer, Haines, and Price (1980). Ten pilots flew a simulated instrument landing approach. The HUD was compared with conventional head-down instrumentation (not collimated). Although most of the landings were normal, on the very last trial, there was a runway incursion. As the pilot was approaching the simulated runway, another aircraft pulled onto the runway. The investigators found that, although the HUD gave better performance under normal landing conditions, a significant number of pilots failed to notice the plane coming onto the runway when flying with the HUD. Furthermore, those that did notice the runway incursion took longer to notice it when they were flying with the HUD than when they were flying with conventional head-down instrumentation. However, this finding was not replicated in a more carefully controlled study by Wickens, Martin-Emerson, and Larish (1993).

The way the NASA investigators interpreted the fixation data was to state that in flying with conventional instrumentation, there is a very regular scan pattern required to check the clearance of the runway; but with the HUD, the imagery may obscure the distant runway, and the scan pattern is disrupted in a way that doesn't allow the routine and automatic examination of the imagery out in the far domain. In the evaluation by Steenblik (1989) of the operational use of the HUD in Alaskan Airlines, some pilots report that in the last few seconds of the approach, coming into and through the flare, they find the imagery on the HUD distracting. They have a tendency to tunnel attention exclusively on that imagery and, therefore, they prefer to turn off the HUD to avoid this tunneling. Also, earlier evaluations done by NASA indicate a substantial problem with tunneling in on the HUD instrumentation and potentially ignoring the outside world. Finally, some research on military applications of the HUD done by

Opatek indicated a lot of problems, at least with early HUD designs, that arose from them being too cluttered, so that pilots had a tendency to turn them off.

A summary of the attentional issues highlights the following points. First, the distinction between conformal and nonconformal symbology is critical. Conformal symbology will not create clutter and clearly is desirable to be presented head-up, particularly when driven by inertial guidance information. Nonconformal symbology, whether digital or analog, may lead to clutter and confusion, and its addition to a HUD, while reducing scanning, should be considered only with caution. Secondly, attentional tunneling on either conformal or nonconformal symbology, to the exclusion of attention to the far domain, is a potentially real problem. Consideration should be given as to how to "break through" the tunnel (e.g., by turning off the HUD or reducing its intensity). Third, there is some suggestion that the tunneling problem may be exacerbated in head-up rather than head-down presentations.

In conclusion, there has been some debate in the aviation psychology literature regarding whether the HUD is an advancement or a detriment to aviation safety. One way of addressing this debate is to point to the strong endorsements provided by pilots who have flown with the current versions. A second way is to consider what HUD has done. It has pushed the performance envelope of aircraft into a whole new domain, and clearly in that new domain there are going to be more chances for risk and accidents, for example, flying lower to the ground in low to zero visibility. In this sense, it is analogous to headlights which, by allowing night driving, have placed the driver in a consistently more dangerous environment (Weintraub & Ensing, 1992).

Chapter 7

Decision Making

by Christopher D. Wickens, Ph.D., University of Illinois

The Decision-Making Process

Figure 7.1 shows a model of information processing. This is similar to the model presented in Chapter 5, Figure 5.1. In the preceding chapters, the discussion focused on basic characteristics of the senses, how the eyes and ears perceive stimuli, and how information from the world around us is perceived or understood. This chapter deals with the decision-making process that takes place after the sensory information is perceived.

Figure 7.1 provides a framework for discussing the decision-making process. A pilot senses a stimulus, for example, the VASI on a runway. That information becomes an understood piece of knowledge when the pilot recognizes the visual

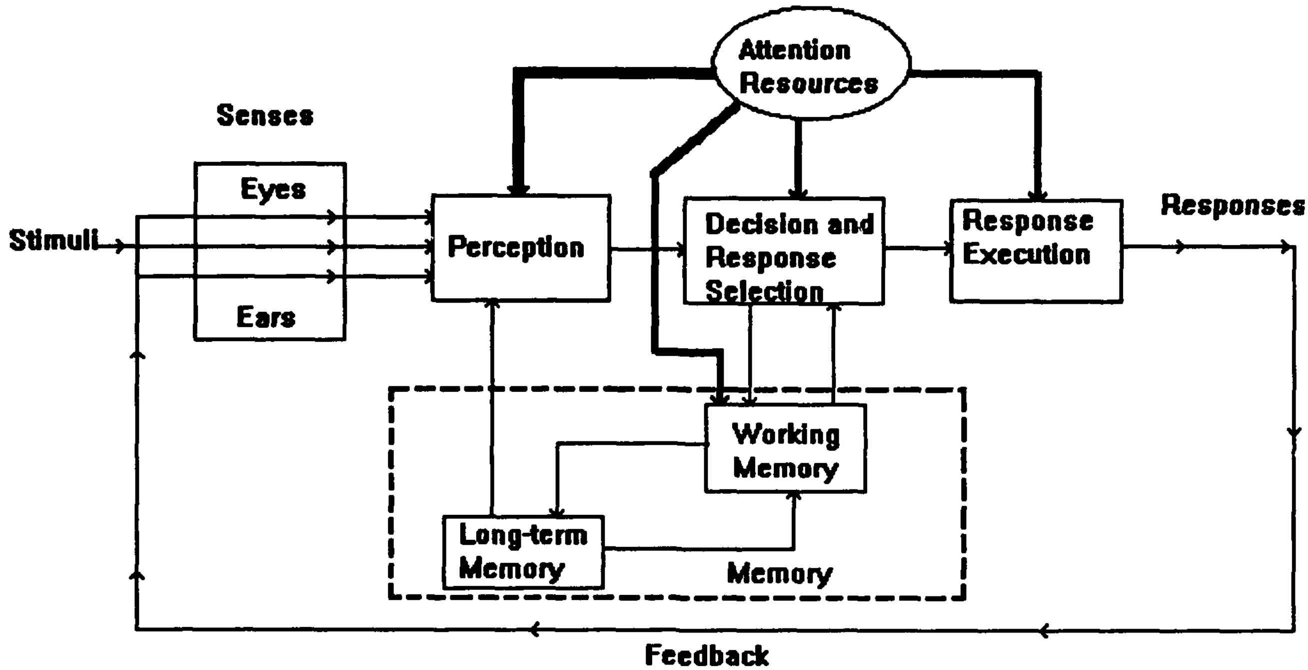


Figure 7.1. A model of information processing. (from Wickens, 1992)

information based upon past experience which is stored as long-term memory. Once perception is complete, then the pilot has a mental representation of the state of things--a situational awareness. He or she is now able to engage in decision and response selection. First, a decision about what to do is made. The decision may be to defer action and hold the information in working memory, or the decision may be how to carry out the response: vocally, manually, by foot movement, etc. After a particular response is selected, the pilot executes the response, that is, carries out some action by coordinated muscular action. The response execution, of course, changes the environment. The new environment provides feedback and new stimuli for the senses, and processing returns to the beginning of the loop shown in Figure 7.1.

Our attentional resources, pictured as a reservoir of limited capacity in Figure 7.1, are critically involved in the decision-making process. Our attention resources are directly applied to perception, working memory, decision/response selection, and response execution. Attention has a limited capacity. It allows us to perceive only so much information at one time, store so much in working memory at once, make only one decision at a time about which responses to execute, and execute so many responses at once. Working memory is particularly subject to the limits of attention resources. Working memory is the very limited capacity buffer where we store temporary information like waypoints, radio frequencies, etc., that we have just received and will

immediately forget if we stop rehearsing. Very often, working memory guides our decisions and responses.

Our discussion will focus on the process of decision making at two levels. First, we consider pilot judgment--the decisions under uncertainty that pilots carry out, generally with considerable thought and effort. Then we consider the rapid and relatively automatic decisions that involve direct selection of an action. This second class has direct relevance to cockpit design issues, and this will lead us to a discussion of the transfer between different designs on different aircraft.

Pilot Judgment

When we talk about decision making, we begin with the concept of uncertainty. Decisions can be made with certainty or with uncertainty. A pilot's decision to lower the landing gear, for example, is made with certainty. The pilot knows he or she must lower the landing gear to touch down on the runway and the consequences of the decision are well known in advance. On the other hand, a decision to proceed with a flight in bad weather or to carry on with a landing where the runway is not visible is a decision with uncertainty, because of the uncertain consequences of the actions. What will happen if the pilot continues with the flight in bad weather can't be predicted.

A lot of the conclusions in decision making that will be discussed here come directly from studies and experiments that have not been related to pilot judgment. There are, of course, databases about aviation accidents and incidents that attribute a large percentage of these to poor pilot judgment and faulty decisions (Jensen, 1977; Nagel, 1988). The problem, of course, is going back after the fact of an accident or incident. It is easy to attribute a particular disaster to poor judgment when, in fact, there may be, and usually are, a lot of other causes. Poor judgment may have been only one of a large number of contributing causes all of which cannot be identified. For this reason, it is helpful to study judgment and decision making in other fields besides aviation, like the nuclear power industry, or to draw inferences from some experimental laboratory research. Much of the information in this section is based upon conclusions from these other nonaviation areas.

Figure 7.2 (Wickens and Flach, 1988) shows a general model of human decision making that highlights the information processing components which are relevant to the decision process. To the left of the figure, we represent the pilot sampling, processing and integrating a number of *cues* or sources of information. If it is a judgment about flying into instrument meteorological conditions, these cues may be weather reports, direct observation of the weather, anecdotal reports from other pilots in the air, etc. All of the cues help

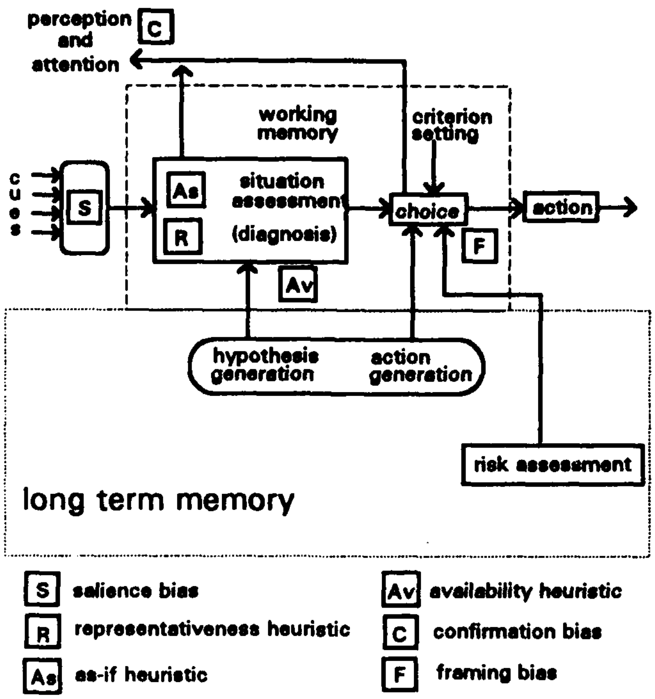


Figure 7.2. A model of human decision making. (from Wickens & Flach, 1988)

the pilot form a ***situation assessment***--what we might call a ***diagnosis***, of what is going on. In making situational assessments, we are often dependent upon our ability to generate hypotheses about what is going on: hypotheses about icing conditions, or severe turbulence for example. Or in the case of diagnosing engine failures, hypotheses about possible failure states of the aircraft. This diagnosis, in turn, depends on the information available from long-term memory, the stored results of training we have had in the past about the things that possibly could go wrong. Having made a temporary situation assessment, we often follow this up by perceiving and attending to further information. In other words, we seek out more cues to either support or refute our hypothesis. So this is very much of a closed-loop process. You form a tentative hypothesis. You go out and get more information, perhaps call for updated weather information, or do more observation to try to confirm the hypothesis. Eventually, you reach a point where a **choice** is required. The choice is between actions also learned and thereby stored in long-term memory. Do you go through with the flight? Do you return to an airport? Do you request an alternate flight path? The choice of an action is sometimes based upon a **criterion setting**, that is, how much information is needed before you carry

through with a given action or decision. In aviation, the criterion setting is very often based upon ***risk assessment***. What is the risk of continuing in bad weather? What is guiding our choice? What are the consequences of failure? And then in the final box in the model in Figure 7.2 we perform an action, and observe its consequences which themselves generate more cues.

Biases in Situation Assessment

The model in Figure 7.2 includes codes (S, R, As, etc.) in small boxes which represent biases that can cause errors in human decision making. Some of these biases are also called ***heuristics***, shortcuts or mental "rules of thumb" that people use to approximate the correct way of making a decision because it takes less mental effort (Kahneman, Slovic, & Tverksy, 1982).

Salience Bias

The first of these biases is called a ***salience bias (S)***. The salience bias means that when someone is forming a hypothesis based on a lot of different cues of perceptual information, he or she tends to pay more attention to the most salient cue. For example, a pilot may be processing various sources of auditory information including weather reports, reports from air traffic control and other pilots, conversation from the first officer, etc., to form a hypothesis. The salience bias is reflected in the fact that **it is often the loudest sound or loudest voice that has the most influence.** Another example of the salience bias occurs in dealing with a multi-element display. **We tend to pay most attention to information displayed at the center of the display rather than the information at the bottom.** These are physical characteristics of a display that aren't necessarily related to how important that information is. The brightness of lights creates a bias: **the brighter the light, the more we tend to pay attention to it in making our situation assessment.**

Confirmation Bias

Early in the decision-making process, we form a tentative hypothesis about our situation and we go back to the environment for more cues. At the tentative hypothesis stage, we may experience a second form of bias called the ***confirmation bias (C)***. The confirmation bias states that once a tentative hypothesis is chosen, **we tend to seek and find information to confirm that hypothesis, but we also tend to ignore information that disputes the hypothesis,** information that tells us we are wrong. An example of the confirmation bias at work is airport misidentification. It seldom happens in commercial aviation, but rather frequently in private aviation. The pilot simply approaches or lands at the wrong airport. There is a tendency when the pilot is lost and disoriented to try to interpret the ground information as consistent with the airport that he is

expecting to approach rather than the airport that he is actually approaching, particularly at night. The visual world (i.e., the pattern of runway lights or surrounding features) is distorted in a way that confirms the pilot's expectations.

The confirmation bias is supported by *expectancy*. What you expect to see helps you confirm what you believe your state is. A major concern in private pilot aviation is continued flight into deteriorating weather. This has been documented by some research at Ohio State University (Griffin & Rockwell, 1989). Pilots continue to pay attention to information saying the weather is good if they have initially filed their flight plan under the assumption of good weather. They ignore the contrary evidence that the weather is deteriorating. Misdiagnosis of failure is another area where expectancy reinforces the confirmation bias. We don't have documented aviation examples of this, but in the nuclear industry there are some very definite situations where operators have formed a hypothesis about a failure state in the plant, and then have sought information to confirm that hypothesis and ignored information that says otherwise. The Three Mile Island disaster can be directly attributed to the effect of the confirmation bias. The operators had a hypothesis that the water level in the plant was too high. They continued to process information that confirmed that, and they ignored much of the other information that indicated, in fact, that the pressure was dropping, and the radioactive core was about to be exposed.

Anchoring Heuristic

A heuristic closely related to the confirmation bias is called *anchoring*. The anchoring heuristic states that if there are a couple of hypotheses you might have, you tend to anchor your beliefs to one and ignore information supporting the other. As new information comes in that supports the other hypothesis, the one you have not anchored to, you don't give it much credibility. So your degree of belief in one versus the other hypothesis doesn't change very much. You are open primarily to the information that confirms the hypothesis to which you are anchored. Then if you get one piece of information that supports what you already believe, (you are already anchored to), you give that information a lot more weight. Again, we can use the example of continued flight into bad weather. If you initially believe the weather is good and that is your hypothesis, you are more likely to process new information that says that the weather is indeed good, and ignore information that says it is poor. One might imagine that different pilots have different beliefs about whether a particular aircraft is a good aircraft or a bad aircraft, or an aircraft system has faults or works well. Biased with these beliefs, the pilot is likely to pay a lot of

attention to information that confirms those hypotheses, and ignore information that doesn't. So if you believe a system works well, you are likely to pay less attention to incidents where the system fails. You are also less likely to notice when the system does fail. If you believe the system is faulty, you are going to be very sensitive to instances in which the system does indeed fail. You may also assume the system has failed when it is, in fact, operating correctly.

Base Rate of Probability

One of the fundamental theories of situation assessment is known as ***Bayes theorem***. Expressed intuitively, Bayes theorem says that whenever you are trying to evaluate or form a hypothesis about what is going on, your belief in the most likely state of the world should be based upon an equal consideration of two things. One is the probability of each state of the world. We call that the ***base rate***. Independent of what you see, how likely is it that the weather is going to be bad versus good? Independent of what you see, how likely is it that your hydraulic system will fail rather than some other failure. In addition to the base rate, the second thing is the **similarity** of the actual data (the available visual or auditory information) to the mental representation of the pattern of symptoms caused by that particular failure. Do the symptoms you see match the pattern of symptoms expected for a given failure?

Bayes theorem can be summarized by the following equation:

$$\text{Belief} = (\text{B} \times \text{Base Rate}) + (\text{S} \times \text{Similarity})$$

Here's an example. You're viewing a particular state of meteorological information. You are trying to form one of two hypotheses: the weather is going to be bad on the route which you are flying or the weather is going to be good. The hypothesis formation should be based upon the similarity between the actual weather that you are viewing and the weather conditions when it is good or bad, and upon the base rate: the probability that the weather indeed will be bad versus the probability that the weather will indeed be good along your route. For example, the base rate probability may be the overall actuarial data that says that at a given location the weather is going to be clear 90 percent of the time, on a given day of the year.

The two elements in Bayes theorem, base rate and similarity, should compensate for each other. So if you don't have much data on which to base similarity, (you haven't got a good weather report and maybe you don't have very good observation of the weather), you should pay most attention to the base rate in making your forecast. That is, what the overall probability is that there will be good or bad weather along your route. On the other hand, if you don't have

base rate data, (if you don't know what those overall probabilities are), and you have a lot of weather forecasts and a lot of good observations, you should pay more attention to the degree of similarity between the hypothesis and the existing conditions.

Availability Heuristic

There are two very important heuristics that we use to approximate the base rate and the similarity of the data to the hypothesis. These are ***availability*** and ***representativeness, respectively.*** We approximate the base rate, how frequent or how probable a certain condition is, by the ***availability heuristic.*** **The availability heuristic leads us to consider a hypothesis most likely if it is most available in memory.** Your estimated base rate of a hypothesis or of a particular risk is based on how easily you can recall that hypothesis from memory. For example, suppose you are trying to diagnose a particular failure state in an aircraft. How probable is it that the failure state exists? There is probably data somewhere about the likelihood that a given system will fail. There is certainly data in the nuclear industry about the probability that certain systems will fail and that data is what you really ought to go on. However, the availability in your memory is governed heavily by ***recency***, by how fresh the information is in your mind. So, according to the availability heuristic, **if you recently experienced a certain kind of failure or perhaps you read about it (in an FAA, company, or other aviation publication) that makes it very available in your memory and, therefore, that failure will seem highly probable.**

In many domains, availability is very much based on ***publicity***. For the flying public, there is a greatly elevated fear or estimation of the probability of a fatal aircraft crash simply because of the high publicity given to aircraft accidents. Because they are very highly publicized, the public generally has available this idea that aircraft accidents are fairly frequent and, therefore, overestimates how likely they are to occur.

Availability is also often governed by ***simplicity***. It is easier to remember or to think about simple situations than complex situations. And this is very much true in trying to diagnose a failure. Multiple failures are fairly complex; therefore, it is hard for people in doing failure diagnosis to think that those multiple failures could happen, because they are simply not easy to recall from memory. It is much easier to think about simple failures; a single element failure rather than a complex failure.

Representativeness Heuristic

We have said that people should rest their belief in part on the base rate probability. We have also said that the way people actually use base rate probability is not by the true probability, but by how easily they can recall instances of an event. However, it seems that people frequently do not use probability at all in making diagnoses. Instead, they attend **only to the similarity or representativeness of the current evidence** or data to one hypothesis or another. The representativeness heuristic further states that the only time we use base rate is when there isn't much data to go on. For example, a pilot may be flying in a particular area, and it is highly probable that the local weather conditions may be severe, based upon past data. If the present weather actually looks clear outside the cockpit, the pilot would tend to ignore the base rate information which might state that in this particular region, at this particular time of year, the weather is likely to degrade. The representativeness heuristic also makes us tend to ignore differences in the probability of different failure states if a set of symptoms that you observe looks like the prototypical case of a particular failure you have well represented in memory. In the case of landing at the wrong airport, the representativeness heuristic would make you ignore the fact that this is really not a likely place for your target airport to be, because the airport runway and the pattern of lights look like the airport you think you should be approaching. The wrong runway is representative of an image you have in memory of the correct runway.

Overconfidence Bias

In understanding where you are, what your situation is, and what you should do next, the ***overconfidence bias*** can be at work. This seems to be a fairly pervasive bias that underlies performance of both novices and experts in a lot of different domains. **We tend to be overconfident in our own judgments** based on our own memory and our own cognitive ability. In other words, when we have solved a problem, we are more confident than we should be that the problem is solved correctly. One important example of overconfidence occurs in eyewitness testimony. A lot of data coming from research on judicial procedures indicate that eyewitnesses to a crime, or to a significant event such as an aircraft accident, tend to be far more overconfident about what they saw than the accuracy of their own testimony will reflect. An eyewitness, for example, might state with high confidence that a plane was on fire before it crashed, when, in fact, it was not. The point is that you can't give much credibility to the eyewitnesses' asserted confidence of what they saw or heard, and instead you must down-weight that confidence appropriately. There has also been some laboratory work done with pilots' decision making at the University of Illinois

where it was found that pilots are more confident that their judgments are correct than they really have a right to be (Wickens et al, 1988).

For the pilot, the consequence of overconfidence in the correctness of a decision that he or she has just made, is that the next course of action will be taken without adequately considering the alternative actions, should the decision in fact be the wrong one, and will be taken without adequately monitoring the evolving consequences of the decision just made.

Risk Assessment

A characteristic of many judgments both on the ground and in the air is the need to choose between a ***risky option*** and a ***sure thing option.*** A risky option has two possible outcomes, neither one of them assured. A sure thing option has only one, certain, outcome. It is almost guaranteed. The classic example of choosing between a risky option and a sure thing option is delaying takeoff on a flight. The sure thing option is that you are going to sit on the ground for a long period of time and nothing is going to happen except a certain delayed flight. The risky option involves going ahead with the takeoff into potentially uncertain weather, a decision with two possible outcomes: an accident or incident due to severe weather, or a safe trip. With the sure thing option, staying on the ground, it is highly probable that everything will be fine, and the consequences of the decision will be generally good (safe, but with a delay). The risky option really has a relatively high probability that things will go very well (a safe flight but no delay), but a very severe negative consequence if the bad weather leads to disaster.

How do people make these choices? Do they tend to go for the sure thing or the risky option? These sorts of decision problems can be expressed intuitively in terms of gambling choices. Here's the choice: you can receive five dollars guaranteed, or you can flip a coin and either win ten dollars or nothing at all. This is really a choice between two positive outcomes with the same expected value in the long run. One is keeping the five dollars, a sure thing. The other is that you have a 50/50 chance of getting something good, ten dollars, or nothing at all. With either option, you have everything to gain and nothing to lose. In contrast, we can also represent these two decision choices in terms of negative outcomes. So I can say, I will take five dollars from you, or you can flip a coin and have a 50/50 chance of either losing ten dollars or nothing at all.

The research in psychology has studied people confronted with these gambling choices (including also trained business people making investments). The results reveal that whether people choose the risky option or the sure thing option,

depends upon whether the choice is between two positive outcomes as in the first example, or two negative outcomes as in the second example. **Given a choice between two positive outcomes, people tend to take the sure thing.** They tend to be aversive to risk, and the expression goes, they "take the money and run." So more people would be likely to take the five dollars than to take the bet of getting more or nothing at all. But **given the choice between two negative outcomes, people tend to be risk-seeking.** The expression for them is, they "throw good money after bad." They are more likely to take the gamble and hope they come out with no loss rather than accepting a guaranteed loss. This difference in choice preference is called ***framing of decisions***, because the way in which a decision is made depends on how it is framed: Whether it is a choice between positives or a choice between negatives (Kahneman, Slovic, & Tversky, 1982).

Consider, for example, a physician making choices between a sure thing medical treatment and risky treatment. Investigations have found that the physician recommendations are very much influenced by whether words are phrased in terms of saving the patient, or the probability that the patient will die. Saving the patient is the positive outcome; the probability of death is the negative outcome.

How do we translate framing into an *aviation-relevant example?* Again, let's consider a decision between, say, canceling or delaying a flight and taking off into bad weather. We can talk about the sure thing characteristics of delaying or canceling the flight. There is a certain good characteristic to delay or cancellation, and that is you are guaranteeing safety. A certain bad characteristic is that you are guaranteeing a lot of irate passengers, a disrupted crew schedule, etc. The risky option of flying into bad weather has a good (but uncertain) outcome: it is likely that you are going to proceed in a more timely fashion. It also has a potentially bad characteristic: with a low probability, it could happen that there is going to be severe delay and possibly disaster. The issue here is that the bias towards one choice or the other can be based on the way in which the positive outcomes are framed or emphasized. Say the decision is between guaranteeing a safe flight or a high probability of getting a timely flight to the destination. That is a decision framed in terms of a positive sure thing and a positive risk. The framing bias suggests that under these circumstances, the bias would be towards delaying the flight and just staying on the ground. Whereas, if the decision were framed in terms of negatives, a sure thing of delay with irate passengers or a relatively small possibility of a crisis because of being in the air in bad weather, there would be a greater bias towards choosing the risky option.

Stress and Decision Making

It is important to consider some of the ways in which stress amplifies the various biases, or otherwise affects decision making. These conclusions are based on both accident and incident reports as well as some experimental data. By stress, we mean the perception of being in a highly dangerous environment, and refer to the kind of experiences that result when alarms start to go off and the cockpit systems start to fail or when the aircraft encounters very serious meteorological conditions. **The effects of stress seem to enhance the confirmation bias, also called cognitive tunneling.** This occurs when you continue to believe in the hypothesis that you initially formulated, regardless of what the new data say. The analysis of the Three Mile Island nuclear incident shows very graphically how the operators, under the stress of a crisis situation after the initial alarms sounded, and knowing they had a critical situation, continued to tunnel in and focus on that one belief that the water level was too high, not too low. **In cognitive tunneling, one not only tunnels to a particular hypothesis, but also tends to focus onto particular elements of a display under high levels of stress and, therefore, process less information.** It is as if the searchlight of attention narrows down onto certain critical cues; you pay most attention to those you believe are most important and you tend to ignore other information. Cognitive tunneling and display tunneling work very much hand in hand, in the sense that the higher the stress the more you pay attention to the information that confirms the hypothesis you believe to be the case. A recent study of errors made by RAF pilots indicated that cognitive tunneling of displays under stress was a significant cause of the accidents they examined. Approximately 16 or 17 percent of the accidents were related to this (Allnut, 1987).

Stress contributes to a loss in working memory: the ability to rehearse digits, (navigational waypoints, radio frequencies), and the ability also to form a mental model of the visual airspace. Research has been done at Illinois that indicates that these imaging capabilities seem to go down under high levels of stress as well (Wickens et al, 1988). Clearly, the more we are stressed, the less we use working memory, and the more we try to use very simple heuristics, simple mental rules of thumb. Under stress, the heuristics or biases tend to dominate our decision-making process.

It is also important to point out that at least some data indicate that there are processes that are ***stress-resistant***. In particular, a lot of times decisions can be made, not by going through this process of weighing all of the information and integrating it with mental calculations, but rather by direct long-term memory retrieval. Decision making by expert pilots in familiar situations is usually automatic and almost unconscious. The pilot sees a situation, it matches

something he or she has experienced before, and that's the diagnosis. The pilot has carried out an action that worked before under those same conditions, so the pilot carries it out again and doesn't go through a time-consuming process of risk evaluation and calculated action choice.

Both the research at Illinois (Wickens et al, 1988) and a lot of the research that Klein (1989) has done with tank crew commanders and with fire fighters indicate that **this type of decision making seems to be much more resistant to stress.** Finally, it has been found that people's ability to evaluate the risk of different options, again, does not appear to be degraded by stress. There isn't a tendency to be more risky or less risky under stress.

Lessening Bias in Decision Making

So where does all this lead to? What steps can be taken to address bias problems in decision making? Clearly, training and developing expertise is one step. Experts tend to use decision strategies that are based more on directly and rapidly retrieving the right action or diagnosis from long-term memory, on the basis of similarity with past experience, rather than using working memory to generate or ponder the alternatives in an effortful manner (Klein, 1989). Another step that can be taken is ***de-biasing.*** There has been some successful research in de-biasing, that is, making pilots or decision makers aware of the kind of biases already mentioned in this chapter. Weather forecasters, for example, if given explicit training about the tendency to be overconfident in their forecasts, can learn to calibrate those forecasts quite accurately. Planning, that is, rehearsing alternatives in advance of a crisis situation, is another step in addressing the bias problem. Effective pilot training naturally strives to get the student to plan for alternative courses of action, and their consequences in different and possible circumstances. Finally one of the more controversial means used to deal with bias, one that is emerging in the commercial flight deck and is already used in the military flight deck, are ***expert systems.*** Expert systems can, at least according to some scientists, replace some of the pilot decision making necessary in the cockpit, or at least can recommend judgments to the pilot in the cockpit.

The mention of expert systems leads us directly to consider the advantages and limitations of automation, an issue covered more thoroughly in Chapter 9. By and large, **automated systems are far more helpful at this stage if they can provide sufficient ways of integrating and presenting information rather than actually replacing judgment and decision making.** There is too little known about the way in which pilots make decisions to trust all of those decision recommendations to the expert systems, but there is much to be gained from using automation to integrate and present information.

High Speed Decision Making: The Choice of Action

A decision to take off rather than abort is one that is made 99 times out of 100, or maybe more frequently than that, without a lot of conscious thought; there is very little choice or uncertainty about the consequence of doing one versus the other. The decision of what key to press on a control display unit is also one you don't really have to think about. You know the consequences of hitting the right keys are good and the consequences of hitting the wrong keys are bad. The decision to respond to a TCAS advisory to engage in a particular flight maneuver is also made without a lot of thought. One doesn't weigh the consequences, (expected cost and benefits), of doing one thing versus another. These are all examples of decision making under certainty. There are many factors that affect how quickly aviators respond to TCAS commands, etc. An important thing to keep in mind as we discuss these factors is that **almost anything that makes a decision take longer will also be more likely to make that decision incorrect. The things that prolong response time are also the same things that will lead to an increased likelihood of error.** (The one exception is the person's choice to proceed more cautiously. The longer decision will probably be more accurate.)

Decision Complexity

The first factor that affects response selection speed is the ***decision complexity.*** The complexity of a decision is literally the number of possible alternatives. Think of a two-choice decision. You are accelerating for takeoff. Do you rotate or abort the landing? There are two possible choices available. A more complex example is a choice between four possible alternatives. A TCAS warning might tell you to turn right or left, or to climb or descend. It might even present more detailed choices: right and descend, left and descend, etc. **The response time increases with the number of possible response alternatives.** In fact, we have a nice equation that can be used to express how long the response time will be as a function of the number of possible alternatives that are available.

$$RT = a + b \log_2 N$$

You can plot this function to show that each time we double those alternatives, we get a constant increase in response time (and an increase in the probability of making a mistake). Again, simple choices are easier and made more rapidly than complex ones.

Expectancy

A second important factor in response time is ***probability***, or expectancy. **We tend to perceive and respond very fast to things we expect, take a long time to respond to (or not perceive at all) things that we do not expect.** For example, in accelerating for takeoff, the pilot very much expects the conditions to be favorable to rotating and going through with the takeoff. He does not expect conditions that will force an abandonment of takeoff procedures. Coming in for a landing, the pilot expects a clear and open runway, and does not expect an obstacle to appear on the runway. We have a formula for the effect of expectancy or probability on reaction time (Hyman, 1953).

$$RT = a + b\log_2[1/p(a)]$$

The lower the probability of the event (a), the less frequent it is, and the longer is the reaction time. These equations provide some evidence, which psychologists are always seeking, for fairly well-defined mathematical laws of human performance. To some extent and in some circumstances, these laws can be balanced against the very strong mathematical laws of engineering performance.

Context

A third factor influencing response selection speed is the ***context*** in which an event occurs. We respond very rapidly if the context makes the event likely. **We respond more slowly if the context makes the event unlikely than if the context makes the event a probable one.** So a crew will respond to a windshear alert quite rapidly if it is in the context of flight into very turbulent thunderstorm conditions near the ground. The crew will respond to the wind shear alert much more slowly if it is in the context of a clear air approach where the weather is good and there is no prior evidence that it is a likely condition to occur. Similarly, the response to a collision or potential collision will be relatively fast in a very dense airspace and will be relatively slow if there is minimal traffic because the latter context is not one that suggests you're likely to encounter traffic.

The Speed-Accuracy Trade-off

Response selection speed is also affected by other factors. The first is a very intuitive one, speed stress. **The more we are stressed to go fast, the faster we go, but the more likely we are to make errors. This is called the *speed-accuracy trade-off*.** On the other hand, the more we try to be accurate in our responses, the slower we are going to be. You can sum this up by saying the higher the accuracy, the higher the response time. A pilot, rushing through a checklist to

reach the next phase of flight as soon as possible, will be more likely to make an error. There is an interesting application of the speed-accuracy trade-off in nuclear power plant design. Designers have found that operators, under emergency conditions, tend to put a self-imposed time stress on themselves. When warning signals start to go off, they want to respond very rapidly. The consequences have been, in a couple of accidents, that people respond fast and make mistakes. Of course, mistakes in dealing with a crisis are the last things you want to have happen. There have been some implicit recommendations in this country, and explicit recommendations in Germany in the nuclear industry, to tell operators that when something starts to go wrong, not to respond immediately. Germany has actually given them a time in which they cannot do anything until they form an understanding of exactly what is happening. In other words, control room operators have been given instructions that combat the tendency to respond fast and make more errors in times of crisis.

Signal and Response Discriminability

Low discriminability between signals is another factor that slows response selection speed. For example, when a pilot is responding to signals rapidly, the likelihood of confusion is greater if the signals are similar to each other. Consider the air traffic controller, for example, who must respond to one of two aircraft that have similar designations, for example, B4723 and B4724. The only difference is the single digit at the end -- 3 and 4, and the controller will take a relatively longer time to respond in this case. On the other hand, if we take away all of those common features and leave only the different features, 3 and 4, the response time can be relatively rapid. Another example of potential discriminability problems might be digital information on head-up displays. If very similar information like air speed and altitude is displayed digitally in a common format, then the high degree of similarity between the representation of each of these may seriously impede a pilot's ability to respond rapidly to a change in one or the other. Auditory alerting tones are also another major culprit for similarity induced slowing or confusion, if there are several different tones, each with different meanings.

Just as two highly similar signals can be confusing and slow down response time to one or the other, so also highly similar switches that have to be used in similar fashion can delay response. If there are two switches that function exactly the same way for different purposes, it will take a pilot longer to pick the right one in an emergency. There is a book called *The Psychology of Everyday Things (Norman,* 1988). It is very readable, nontechnical, and it demonstrates how the selection of action is influenced by the design of everyday things like automobile dashboards, VCR controls, light switch controls, etc. For example, it discusses the problems associated with clock-radios having

what the manufacturer calls "human-engineered" direct input pushbutton control, in which all of the controls look identical. This is exactly the opposite of good human engineering principles where you would want to have a high degree of discriminability between one control and another.

Practice

Practice is still another influence on response time. **The more practiced we are at responding in certain ways under specific conditions, the more rapidly those responses will be.**

The Decision Complexity Advantage

We have already seen that complex choices take longer than simple choices. A four-choice reaction takes longer than a two-choice reaction. However, there are situations in which it is better to have a smaller number of complex choices than a large number of simple choices. And that we call the ***decision complexity advantage***. A good example of this is going through a menu on a flight management computer. What does a pilot need to do if he goes through a menu? There may be a total of 16 options, one of which has to be selected. How do you get to those 16 options to choose the one you want? One possibility is to put all 16 options on a single menu page and have the pilot choose from the 16 items. This is called a "broad/shallow" menu and involves one complex decision. Another option might be to break them into four groups of four options, and have the pilot first choose which group of four he wants to use. Then once he gets the group of four options, he makes another choice within the remaining four. This is called a "narrow deep" menu and involves two simpler decisions. The suggestion is that there are a lot of different ways of getting down from the beginning of a menu to the final option you want.

So which is better: broad shallow menus, with lots of options/menu, or narrow deep menus? **It is generally better to make a smaller number of more complex choices (broad/shallow) than a larger number of slightly simpler choices (narrow/deep).** That is a fairly well-established principle in human factor design (Wickens, 1992). This is some of the kind of guidance that human factors is able to offer for that issue.

Following Checklist Procedures

Menu choice is one case where operators have to execute a number of responses in sequence. Another area in which multiple responses are relevant is in following checklist procedures; a topic that is well discussed from a human factors viewpoint by Degani & Wiener (1990). One of the greatest potential causes of human error is in following a checklist. Here again, there are some

human factor guidelines that are relevant. **A major point in checklist design is to avoid negatives.** Negatives in any sort of checklist or procedural instruction do two things. First of all, they provide an added cognitive load. Any time you hear or read "do not do something" you have to mentally represent what it is that you do and then mentally reverse that representation. Psychologists have shown that this added cognitive transformation takes longer, and it also increases mental workload. A second danger of negatives in a checklist is that there is always the possibility of missing the negative and assuming that it is the positive. Negatives are particularly dangerous in command information. When someone needs to know what should be done, the instruction should be a positive one. If a pilot should ascend, it is confusing to command "don't descend," but saying "ascend" or "climb" is clear. Negatives should also be avoided in communicating status information. To be told that the status of something is "not" this or that places that extra burden on the mind to translate the negative information to a positive. Also, saying what something is "not" is ambiguous, because there are often several things that it could be instead.

Another important issue related to checklists is the idea of ***congruence***. Anytime there is a checklist or verbal narrative of things to be done that will be played out, in some sequence, over time, **it is important to make sure that the ordering of words over time is congruent with the ordering of time.** If you are reading a checklist that says "do X then do Y," you encounter the letter X before Y, and that is the correct, congruent order. If you have a checklist that says "before you do Y, make sure X is done," then you have an ordering of the words in the checklist that is opposite or incongruent from the ordering of actions that are to be accomplished. So somebody who quickly glances at the sequence sees Y then X, and perhaps does Y first, which is reversed from the intended order.

There is a lot to be gained by the use of pictures in checklists and procedure following. Here we are getting more into the maintenance guidelines rather than the flight deck guidelines, but it is still relevant. Figure 7.3 includes an instruction written in text and the same instruction illustrated with a picture and text. The text-only verbal instructions are "See that the sliding cog associated with the reverse drive bevel is rotating freely before tightening the long differential casing." A better presentation is the drawing combined with brief instructions, numbered (1) and (2). This is a clear case of where a picture speaks far more clearly than words. Another characteristic of this picture that might be considered in terms of the logical way of processing information is that we typically read from left to right. Therefore, following the principle of congruence, it would be better to have instruction (1) to the left of (2), so you encounter instruction (1) first as you scan from left to right.

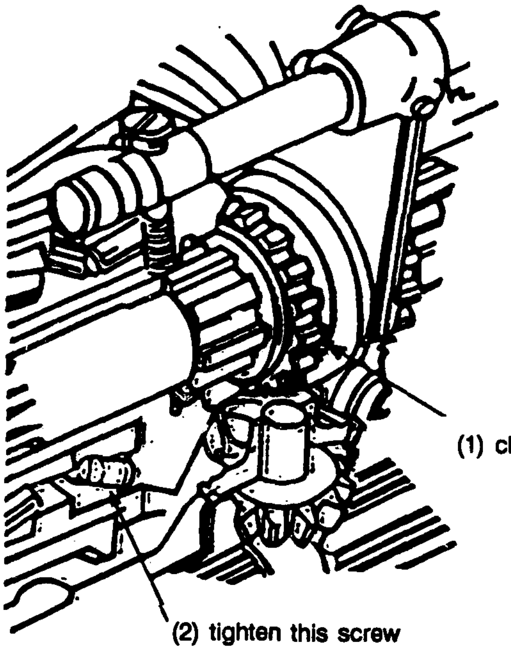


Figure 7.3. Example of how an illustration can be used to avoid technical jargon and improve comprehension. (from Wright, 1977)

Response Feedback

Another issue in response selection, particularly relevant to making several responses in a row, is the issue of feedback from the responses. There are two different classes of feedback. ***Extrinsic feedback*** is separate from the act of making the response itself. Extrinsic feedback is often visual. For example, when you press a key on a CDU (control-display unit), you see a visual indicator on the display corresponding with the key that was pressed. ***Intrinsic feedback***, on the other hand, is directly tied to the act itself. It may be tactile feedback where you press a button and feel the click as it makes contact, and perhaps you hear a click. **Intrinsic feedback is very useful if it is immediate; that is, if it occurs immediately after the action.** For example, pushbutton phones that give you a tone each time you press a button provide better intrinsic feedback than those that don't. There is a great advantage to making sure any keyboard design includes this intrinsic, more immediate feedback.

On the other hand, it is clear that **delayed feedback is harmful, particularly for novices.** It disrupts the ability to make sequential responses, particularly when that feedback is attended to and is necessary, or particularly if it is intrinsic. One of the things we have known for a long time is that delayed auditory feedback has a tremendously disruptive effect. If you are hearing your own voice and it is delayed by as little as a quarter second, the voice transmission is very profoundly degraded. Looking toward the future, design considerations for the data-link system between pilots and area traffic controllers will need to be concerned with feedback issues, as the pilot communicates through the computer interface with the ground using various forms of non-natural displays and non-natural controls, (i.e., keyboard controls, computer-based voice recognition, and voice synthesis).

Display-Control (Stimulus-Response) Compatibility

The compatibility between a display and its associated control has two components. One relates to the relative ***location*** of the control and display; the second to how the display reflects (or commands) control ***movement***.
In its most general form, the principle of location compatibility says that the location of a control should correspond to the location of a display. But there are several ways of describing this correspondence. Most directly this is satisfied by the principle of ***colocation***, which says that each display should be located adjacent to its appropriate control. But this is not always possible in cockpit design when the displays themselves may be grouped together. Then the compatibility principle of ***congruence*** takes over, which states that the spatial arrangement of a set of two or more displays should be congruent with the arrangement of their controls. Unfortunately, some aviation systems violate the congruence principle (Hartzell et al., 1980). In the traditional helicopter, for example, the collective, controlled with the left hand, controls altitude which is displayed to the right; whereas the cyclic, controlled by the right hand, affects airspeed which is displayed to the left.

The distinction between "left" and "right" in designing for compatibility can be expressed either in relative terms (the airspeed indicator is to the left of the altitude indicator), or in absolute terms, relative to some prominent axis. This axis may be the body midline (i.e., left hand, right hand), or it may be a prominent axis of symmetry in the aircraft, like that bisecting the ADI on an instrument panel, or that bisecting the cockpit on a twin seat design. Care should be taken that compatibility mappings are violated in neither relative nor absolute terms. For example, in the Kegworth crash in the United Kingdom in 1989, in which pilots shut down the remaining, working (right) engine on a Boeing 737, there is some suggestion that they did so because the diagnostic

indicator (engine vibration) of the malfunctioning (left) engine was positioned to the right of the cockpit centerline (Flight International, 1990).

Sometimes an array of controls (e.g., four throttles) are to be associated with an array of displays (e.g., four engine indicators). Here, congruence can be maintained (or violated) in several ways. Compatibility will be best maintained if the control and display arrays are parallel. It will be reduced if they are orthogonal (Figure 7.4, i.e., a vertical display array with a horizontal left-right or fore-aft control array). But even where there is orthogonality, compatibility can be improved by adhering to two guidelines: (1) the right of a horizontal array should map to the front of a fore-aft array; (2) the display (control) at the end of one array should map to the control (display) at the end of the other array to which it is closest (see Figure 7.4). It should be noted in closing, however, that the association of the top (or bottom) of a vertical array with the right (or high) level of a horizontal array is not strong. Therefore, ordered compatibility effects with orthogonal arrays will not be strong if one of them is vertical. Some other augmenting cue should be used to make sure that the association of each end of the array is clear (e.g., a common color code on both, or a painted line between them).

The movement aspect of SR compatibility is called ***cognitive-response-stimulus compatibility*** or ***CRS-compatibility***. This means that the pilot has a cognitive intention to do something: increase, activate, set an air speed, turn something on, adjust a command altitude, etc. Given that intention, the pilot makes a response, an adjustment. Given that response, some stimulus is displayed as feedback from what has been done. There is a set of rules for this kind of mapping between an intention to respond, a response, and the display stimulus. The rules are based on the idea that, first of all, people generally have a conception of how a quantity is ordered in space. As we noted in the previous chapter, when we think about something increasing, we think about a movement of a display that is either upwards, to the right, forward, or clockwise. Secondly, there is a set of guidelines having to do with the relationship between control and display movement that is most compatible, or that is most natural. These guidelines are shown in Figure 7.5. Whenever one is dealing, for example, with a rotary control, there are certain expectations we have about how the movement of that control will be associated with the corresponding movement of a display. We think of these as stereotypes, and there are three important stereotypes.

The first is the ***clockwise increase stereotype***, meaning anytime we grab a rotary control, if we want to increase the quantity, we automatically think we have to rotate the rotary control in a clockwise direction (c and d). The second stereotype is what is called the ***proximity of movement stereotype***. It says that

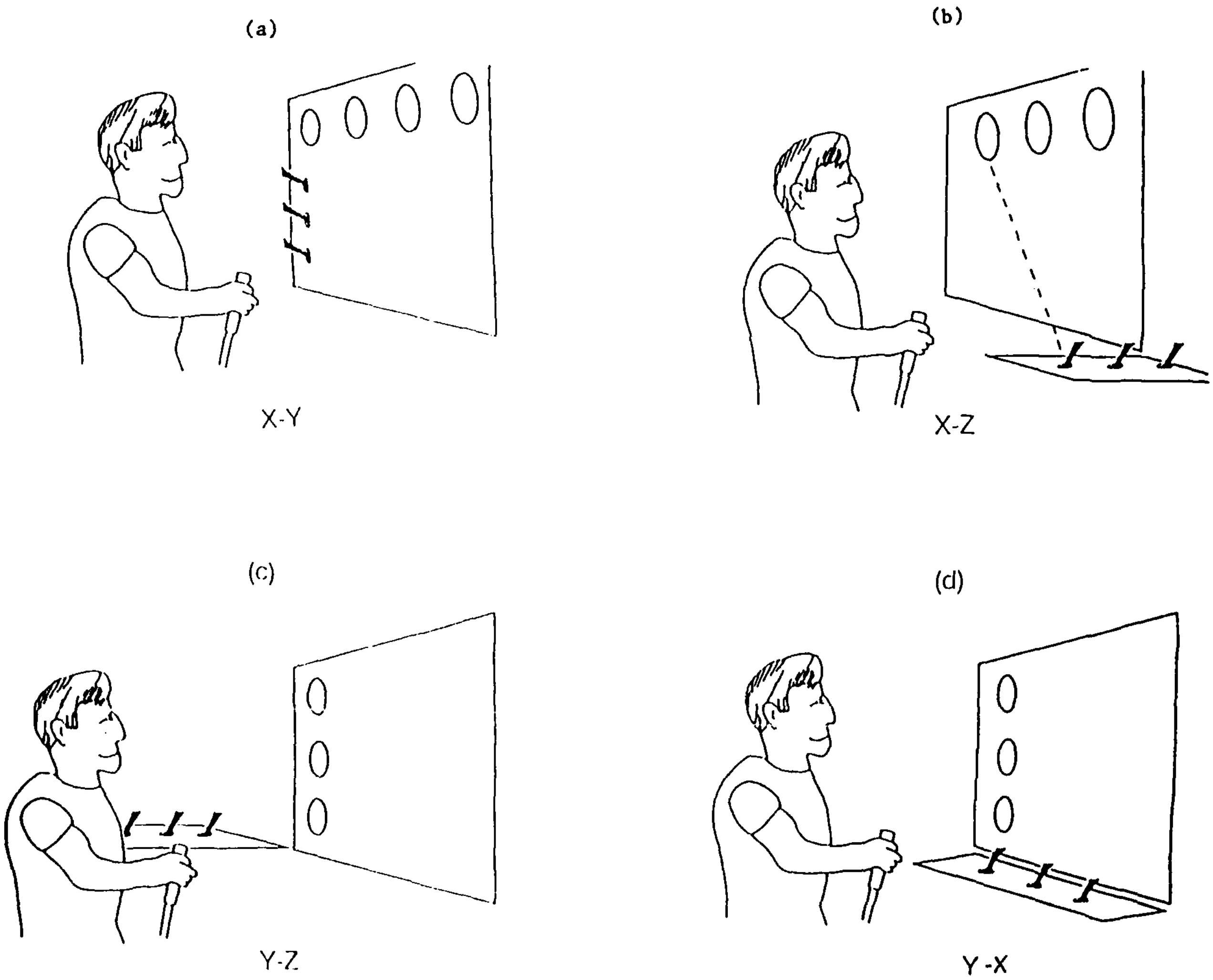


Figure 7.4. Different possible orthogonal display-control configurations. (from Andre & Wickens, 1990)

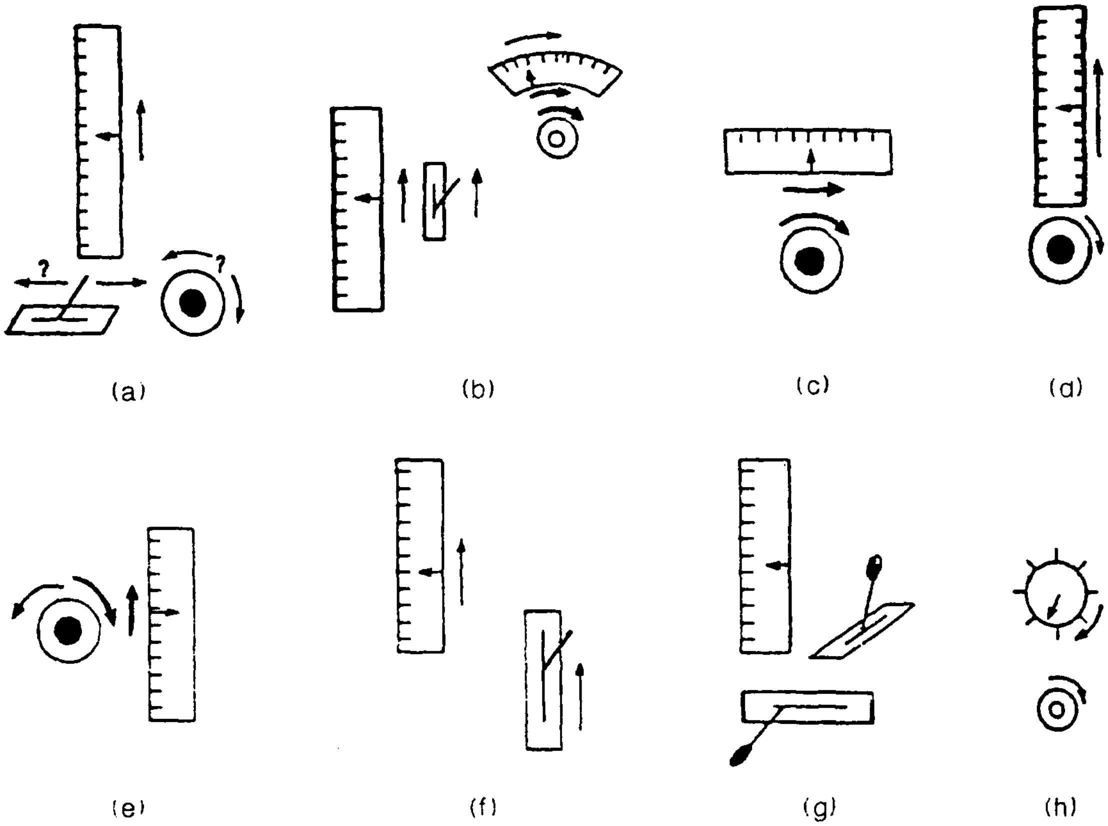


Figure 7.5. Examples of population stereotypes in control relations. (from Wickens, 1988)

with any rotary control, the arc of the rotating element that is closest to the moving display is assumed to move in the same direction as that display. Looking at (c) in Figure 7.5, we see that rotating the control clockwise is assumed to move the needle to the right, while rotating it counterclockwise is assumed to move the needle to the left. It is as if the human's "mental model" is that there is a mechanical linkage between the rotating object and the moving element, even though that mechanical linkage may not really be there.

The important point is that it is very easy to come up with designs of control display relations that conform to one principle and violate another. A good example is (e). It shows a moving vertical scale display with a rotating indicator. If the operator wants to increase the quantity, he or she grabs the dial and rotates it clockwise. That will move the needle on the vertical scale up, thus violating proximity of movement stereotype. You can almost hear the grinding of teeth as one part moves down while the adjacent part moves up. How do we solve the confusion? Simply by putting the rotary control on the right side rather than the left side of a display. We have now created a display control relationship that conforms to both the proximity of movement stereotype as well as the clockwise to increase stereotype. Simply by improving the control-to-display relationship, designers can reduce the sorts of blunder errors that may occur when an operator inadvertently sets out to, say, increase an air speed bug by doing what seems to be compatible, and instead it moves it in the opposite direction.

The third component of movement compatibility relates to congruence. Just as we saw with location compatibility, so movement compatibility is also preserved when controls and displays move in a congruent fashion: linear controls parallel to linear displays [(f), but not (g)], and rotary controls congruent with rotary displays [(b) and (h). Note, however, that (h) violates proximity of movement]. When displays and controls move in orthogonal directions, as in (g), the movement relation between them is ambiguous. Such ambiguity, however, can often be reduced by placing a modest "cant" on either the control or display surface, so that some component of the movement axes are parallel, as shown in Figure 7.6.

As we have seen with the proximity of movement principle, movement compatibility is often tied to a pilot's "mental model" of the quantity being controlled and displayed. Figure 7.7 shows one particular example of display-to-control compatibility that indicates how consideration of the mental model can increase the complexity of compatibility relations. This example is taken from an aircraft manual on a vertical speed window. It is a thumbwheel control mounted in the panel, and to adjust the speed down, you rotate the wheel upward. The label next to the thumbwheel shows an arrow pointing up to

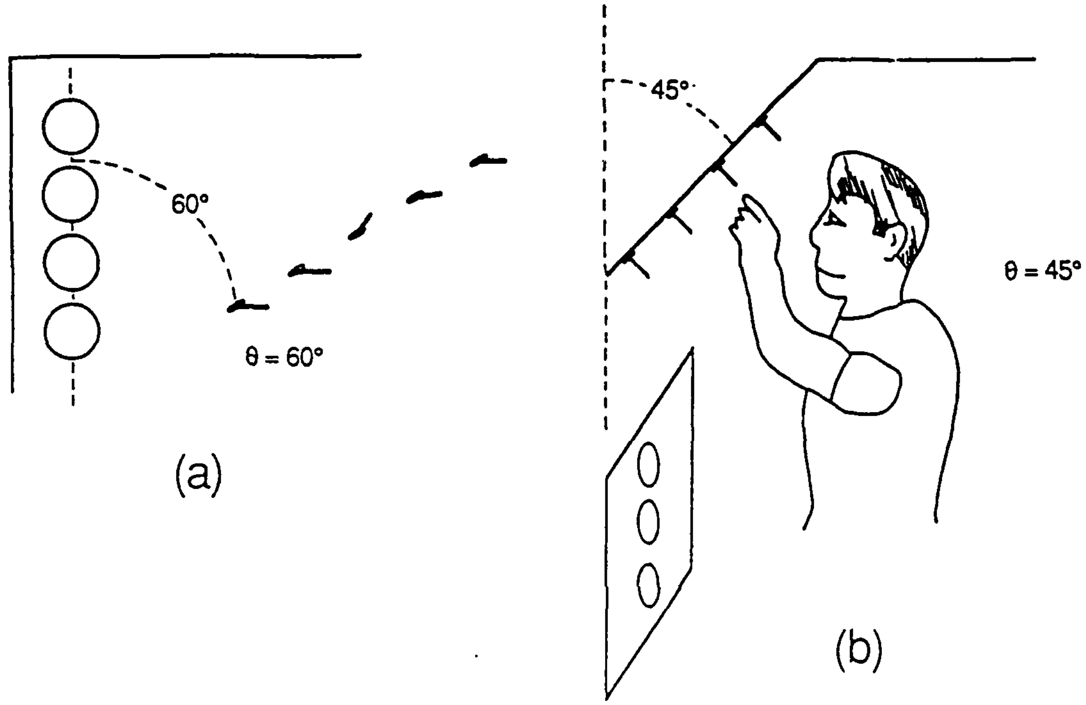


Figure 7.6 Illustration of how a "cant," i.e., angling controls to be partially parallel to displays will reduce compatibility ambiguity. (from Andre & Wickens, 1990)

bring down (DN) vertical speed and an arrow pointing down to bring vertical speed up (UP). From the human factors point of view, this is an incompatible relationship between control and display. If you want to go down, you should push something down, not up. If you want to go up, you should push something up. However, consideration of the mental model makes the relation more compatible than it first appears. If you think about this as a vertical wheel, mounted into the cockpit along the longitudinal axis, you are basically

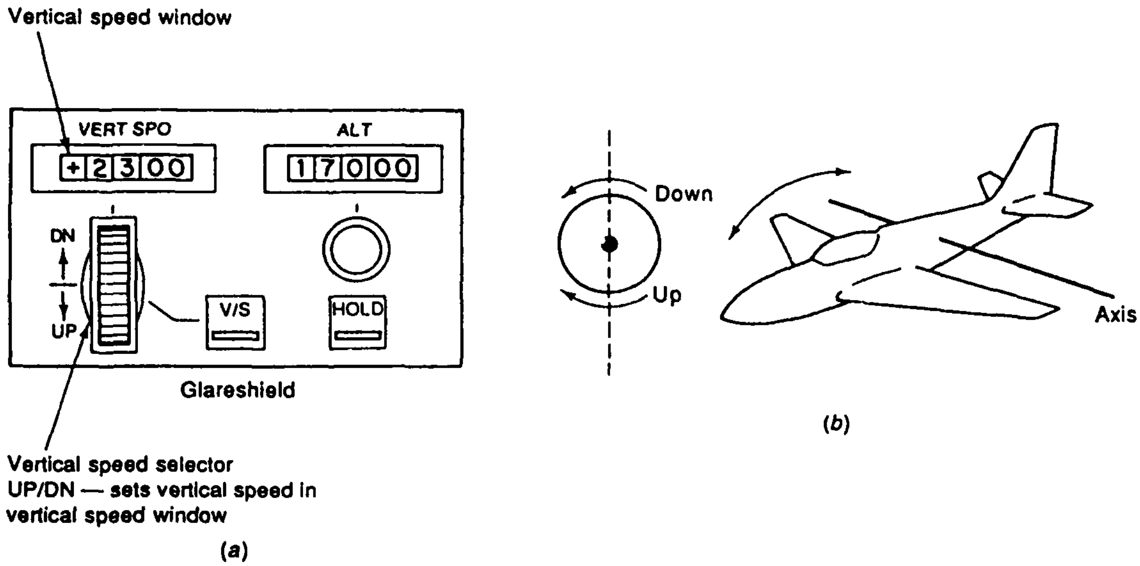


Figure 7.7. Example of display-to-control compatibility on a vertical speed window. (from Wickens, 1992)

rotating the nose of the aircraft down or up. So moving it up rotates the nose of the aircraft down, thereby creating a descent. How pilots think of this is not altogether clear, but it illustrates an important principle that a pilot's mental model of what a control is doing has tremendous implications for whether that control will be activated in the correct or incorrect direction.

Compatibility concerns also address the issue of how a toggle switch should move to activate or provide power to a system. To configure a control mounted on a front panel in a way that its movement will increase the quantity of something or activate it, we might well have it move to the right or upward. If it is mounted along a side panel, we might want to move forward to increase (on) and backward to decrease (off). What happens when we have it mounted on a panel which is at an angle between the right side and the front? We now have a competition between whether this panel is being viewed as closer to the forward position, in which case an increase should be to the right, or closer to the sideward position, in which case an increase should be forward--but in the opposite direction. Which way should this control go to increase? An answer is: Why fight the stereotypes? Why not instead go with the one direction that is unambiguous. That is, make sure upwards increases? If there is a zone of ambiguity, where you have one stereotype fighting against the other stereotype,

good human factors should consider that battle and take advantage of designs that make sure that neither stereotype is violated.

The idea that "on" is indicated by up, right and forward moving switches is contradicted by at least one design philosophy. Figure 7.8 shows the "sweep-on" switch position concept illustrated for a pilot in a cockpit. The sweep-on concept says that to turn switches on, a pilot can do so with a single continuous sweep of the hands. So the direction for on is forward at the bottom, but is backwards up at the top of the cockpit control panel. While there is a certain amount of logic behind this design, given the simplicity of movement, it also presents a concern if a pilot must suddenly focus on a switch overhead and makes a rapid decision whether it is on or off. Does the fact that it is thrown in a backward position counteract the stereotype that means that forward means on? Again, it is not an issue that is easily settled. It is the kind of issue for which a lot more data should be collected to find out how these different stereotypes can come into conflict with each other, and when they do, which one "wins."

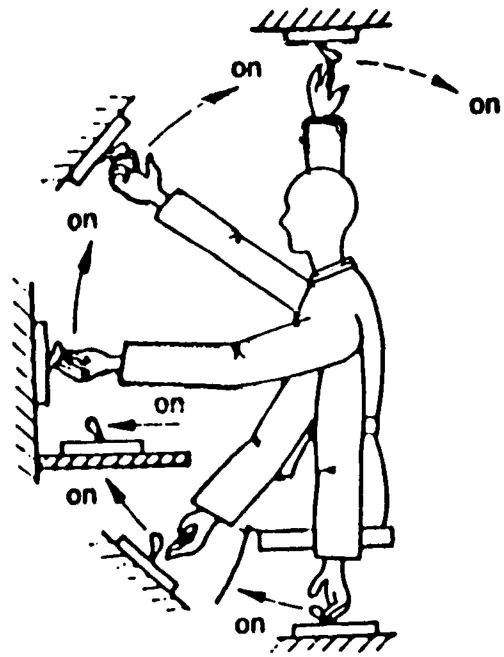


Figure 7.8. The 'sweep-on' switch position concept which is slowly replacing the earlier 'forward-on' arrangement. (from Hawkins, 1987)

SR-compatibility is also related to ***modality***, both voice versus visual display, as well as voice versus manual control. Not a lot of work has been done in this area. We are going to see, and already are seeing in the military, more and more voice-activated controls replacing manual controls. Certain guidelines seem to exist that suggest that voice control is well-suited (compatible) for certain kinds of cognitive tasks, but poorly suited (incompatible) for other kinds of tasks. The voice is very good for making categorical output, describing a state. On the other hand, using the voice for any sort of tracking task, describing the location of things, or movement of things in space, is relatively poor. One reason for this is that our understanding of space is directly connected with our manipulation

of the hands. Therefore, the hands, whether using a key or joystick are much more appropriate for continuous analog control when responding to continuous analog displays. The one possible benefit for voice control of continuous variables would occur if the hands were already heavily involved with other manual control activities. (See Chapter 8.)

Stress and Action Selection

As we have mentioned before, high stress tends to shift one towards fast but inaccurate performance. People tend to react rapidly, but they tend to make more mistakes. It is also clear that under stress, people shift to the most compatible habits and actions. This is probably the strongest reason for keeping stimulus-response compatibility high. Under low stress, people can be effective using an incompatible design like an overhead switch that goes back to turn something on. However, the data suggests that under high levels of stress, the incompatible design is likely to cause an accident, even for the skilled pilot. Somebody wants to turn it off, so by habit they move it backward (which is really on). So compatibility is most beneficial under stress, and, of course, the 1 percent of the time when stress is high is when we are most concerned about good cockpit design, because this is the period in which the environment may be least forgiving of human error.

Stress also has other effects on action selection. It biases operators to perform the best learned habits, in place of more recently learned habits. Stress leads to a sort of "action tunneling," which is analogous to the cognitive tunneling we discussed above. In action tunneling, the pilot may repeat the same (unsuccessful) action over and over. Because stress reduces the capacity of working memory, it may have a particularly degrading effect on ***multimode systems***--like a multimode autopilot--in which the pilot must remember what mode of operation a system is in, in order to select an appropriate action. (We discuss these systems again under the topic of human error in the next chapter.) If the memory fails (because of stress), the multimode system becomes particularly vulnerable to an inappropriate action.

Finally, stress has implications for voice control, where either a pilot or air traffic controller is talking to voice recognition systems. Major concern in the research on voice control is the extent to which high levels of stress distort the voice quality and, therefore, distort the computer's ability to recognize and categorize the voice message. This has been one of the biggest bottlenecks to the use of voice control in military systems. What happens when a pilot comes under stress when talking to the aircraft, and the aircraft does not recognize his voice commands?

Negative Transfer

The topic of stress and action selection are closely related to the issue of ***negative transfer***. Negative transfer is the bringing of habits used in one environment into another environment where those transported habits now conflict with the actions that are called for. There are problems of negative transfer when a pilot transfers from one aircraft to another, when a pilot deals with, say, a modification in his or her customary aircraft, or even when a pilot deals with two different systems within the same aircraft like two different keyboards. Wiener (1988), for example, has called attention to the negative transfer between the ACARS and FMC keyboards in many modern commercial aircraft. The negative transfer issue is directly relevant to the whole issue of pilot certification and ***common type rating***. At what point should two aircraft have different type ratings that require major differences in training?

An example of an accident that was directly related to negative transfer occurred on the DC-9 that crashed on an ILS approach. The new, modified DC-9 involved replacement of the flight director system. In the old system, a full clockwise rotation of the mode selector switch engaged an approach mode. In the new system, the same clockwise rotation of the mode selector engaged a go-around mode. So the same action produced two very different results in the old and new systems. In the analysis of the accident, Rolf Braune reconstructed a sequence in which the crew presumably intended to do an approach, and inadvertently selected a go-around mode by turning the mode selector clockwise. That caused the confusion that led to the accident.

Given potentially catastrophic confusions such as that described above, designers need to be concerned with the causes of a negative transfer, as well as ***positive transfer*** in which experience with the previous system helps performance with the new system. **The most general principle of negative transfer is that unless two designs are identical in both appearance and procedure, the following design changes will increase the potential for crew error:**

- o **The appearance of the new design is the same or similar to the old.**
- o **The procedure is similar, but not exactly the same.**

Table 7.1 is a matrix showing error probability due to transfer of previous learning and experience. Almost any task that a pilot must perform can be characterized by some perceived information read from a display and a required action. This matrix portrays whether the perceived information and the required action is the same between the old and the new systems.

Table 7.1.
Matrix Showing Error Probability Due to Transfer. (from Braune, 1989)

	Perceived Information	Required Action	Transfer of Previous Learning and Experience	Error Probability Due to Transfer
Case 1	Same	Same	Maximum Positive	None
Case 2	Different	Same	Positive	Immediate
Case 3	Different	Different	Little or None	Low
Case 4	Same	Different	Negative	High

In Case 1 in Table 7.1, the perceived information is the same and the required action is the same. With two identical systems, therefore, everything that was learned in the old system is going to transfer to performance in the new system. There is going to be a maximum positive transfer of previous learning and experience from the old system to the new. There is really no possibility for errors in the transfer.

Case 2 is where there is a different representation of the perceived information, but the same required action. For example, the old system might have an analog display and the new system has a digital CRT display. The information is perceived differently because it is presented in two different formats but the required action is the same. The transfer of previous learning and experience will be positive. Error probability is intermediate, so that some errors will occur but not a great many.

In the Case 3 example, both the displays and the controls are different. Therefore, there is little or no transfer of previous learning and experience. The probability of error due to transfer in Case 3 is low. In Case 4, the perceived information is the same, but there is a different required action. This was the situation in the DC-9 crash. The same mode switch in two cockpits performed different actions. The mode switch had to be set differently in the old system than in the new system, and here is where the transfers of previous learning and experience are highly negative. These are the "red flags" for potential error in transferring from one design to the other.

It is important to note that the potential for negative transfer is greatest when the required action is actually similar, but incompatible with the old action. In the DC-9 crash described above for example, the identically appearing rotary switch was turned in both cases; only the turn was to a different position in the old and new (two incompatible responses). The nature of the transfer relationship shown in the matrix is such that negative transfer may sometimes be avoided by making the appearance of the new response device substantially different from the old (e.g., a pushbutton select, rather than a rotary control, in the above case). One of the greatest problems with the different aircraft manufacturers doing their own thing is the extent to which there is a lack of standardization of those kinds of display-action relations across aircraft. In particular, there is a lack of consistency in the relationship between computer systems and control that leads operators to make errors when transferring from one to the other.

Chapter 8

Timesharing, Workload, and Human Error

by Christopher D. Wickens, Ph.D., University of Illinois

Divided Attention and Timesharing

In Chapter 6, we talked about attention in terms of ability to divide attention between two different sources of displayed information. We talk now of attention in the broader sense of being able to divide attention between a large number of different tasks such as between flying and communicating, between navigating and talking, or between understanding the airspace and diagnosing the failure. Discussion of attention in these terms describes issues of **timesharing**. Each of these shall now be described in turn, before addressing the broader issues of workload and human error.

Sampling and Scheduling

The first mechanism relates to task sampling and scheduling; that is, how well does an individual know what perceptual channel or task to attend to at what time. Effective timesharing is being able to attend to the right thing at the right time. Much of your ability to take notes at a lecture is based on your ability to write when the speaker is not saying anything important, then switch your attention to listening when the speaker is saying something important. A lot of research on selective attention, on being able to attend to the right place at the right time, particularly in aviation, has focused on the visual world and pilots' successful ability to look at the right instrument at the right time. The general conclusion of research at NASA Langley is that pilots are fairly good at attending to the right place at the right time.

On the other hand, there is also some good evidence that task scheduling and information sampling is not always optimal. Accident reports may be cited in which pilots have clearly "tunneled" their attention onto tasks of lower priority, while neglecting those of higher priority (e.g., maintaining stability and safe altitude). The Eastern Airlines crash into the Florida Everglades in 1972 is perhaps the most prominent example. Furthermore, experiments done at Illinois find that student pilots do not adequately postpone lower priority tasks when workload becomes high.

There is some interesting research that Gopher (1991) has done with the Israel Air Force which looks at ways to train pilots to better allocate their attention flexibly between tasks. This training device was found to be fairly effective in qualifying pilots for fighter aircraft duty.

Confusion

A second cause of poorly divided attention in doing two things at the same time relates to confusion, a topic discussed in our section on HUDs. You can think of two channels of information, and two responses, but the responses that should have been made for B show up in A, and the responses that should have been made for A show up in B. Recall our discussion of a pilot flying an HUD. There is a motion in the outside runway because the plane changes attitude, and the pilot interprets that motion as being motion on the HUD. This is an example of confusion. One possible way of avoiding confusion between HUD imagery and the far domain is by the use of color. Certainly confusion often occurs in verbally dependent environments where there are two verbal messages arriving at once; for example, a pilot listening to a copilot and simultaneously listening to an air traffic controller. There is confusion when a message coming from one person gets attributed to the other person, or when the digits or the

words in the two messages get confused. The main guideline to avoid confusion is to maximize the differences between the voices. You are less likely to confuse the voice of the copilot with the voice of the controller if one is male and the other is female than if both are male or both are female. The same thing could probably be said regarding digital voice messages. Make sure the voice quality of the digital message is very distinctive and very clear, perhaps by making it sound mechanical, which differs markedly from the voices typically heard on the flight deck. Differences that help us to distinguish between voices include location (or source) and pitch.

Resources

The third mechanism that is involved in timesharing and attention when doing several things at a time is the concept of ***resources***. We have limited capacity, resources, or a supply of "mental effort" that is available for different tasks. Because this limitation exists, the concept of processing resources is important to the issue of pilot workload prediction and assessment, a topic to be discussed later in the chapter. We allocate our limited attentional resources to tasks; as we try to do two tasks at once, for example, fly and communicate, one task gets a certain amount of resources and another task receives the remainder. Our ability to do the two activities at once depends upon the demand of the task for resources and the available supply. In discussing task demand and supply of resources, psychologists describe a function that relates the level of performance on a given task to the amount of resources that are invested in that task. This function is known as the ***performance resource function***. If you take a very difficult task, for example, flying through heavy turbulence and landing under low visibility conditions, it requires a full investment of all of one's resources. One hundred percent of the resources are required to obtain a given level of performance, and that level of performance isn't very good. However, if you consider an easy task, like cruising through clear weather, one can obtain very good performance by only investing half of the attentional resources; and trying harder (investing more resources) can't improve performance any further. You can get maximum performance by giving only a small amount of your resources.

Figure 8.1 presents the performance-resource functions for an easy task (top), a difficult task (bottom), and one of intermediate difficulty. The difference between the bottom and top curve is important not only in the level of performance that is attainable, but also in the amount of "residual resources" that are available to devote to a second (concurrent) task. For the difficult task, as for the intermediate one, any diversion of resources to a secondary task will sacrifice its performance. But for the easy task, a good portion of resources can be diverted with no loss in performance.

The curves in Figure 8.1 are also related to training. Extensive practice on any given task will shift the performance resource function from the bottom, to the middle, to the top curve. As the task can be performed with fewer resources, we say that its performance has become ***automatized***. Compare the middle and top

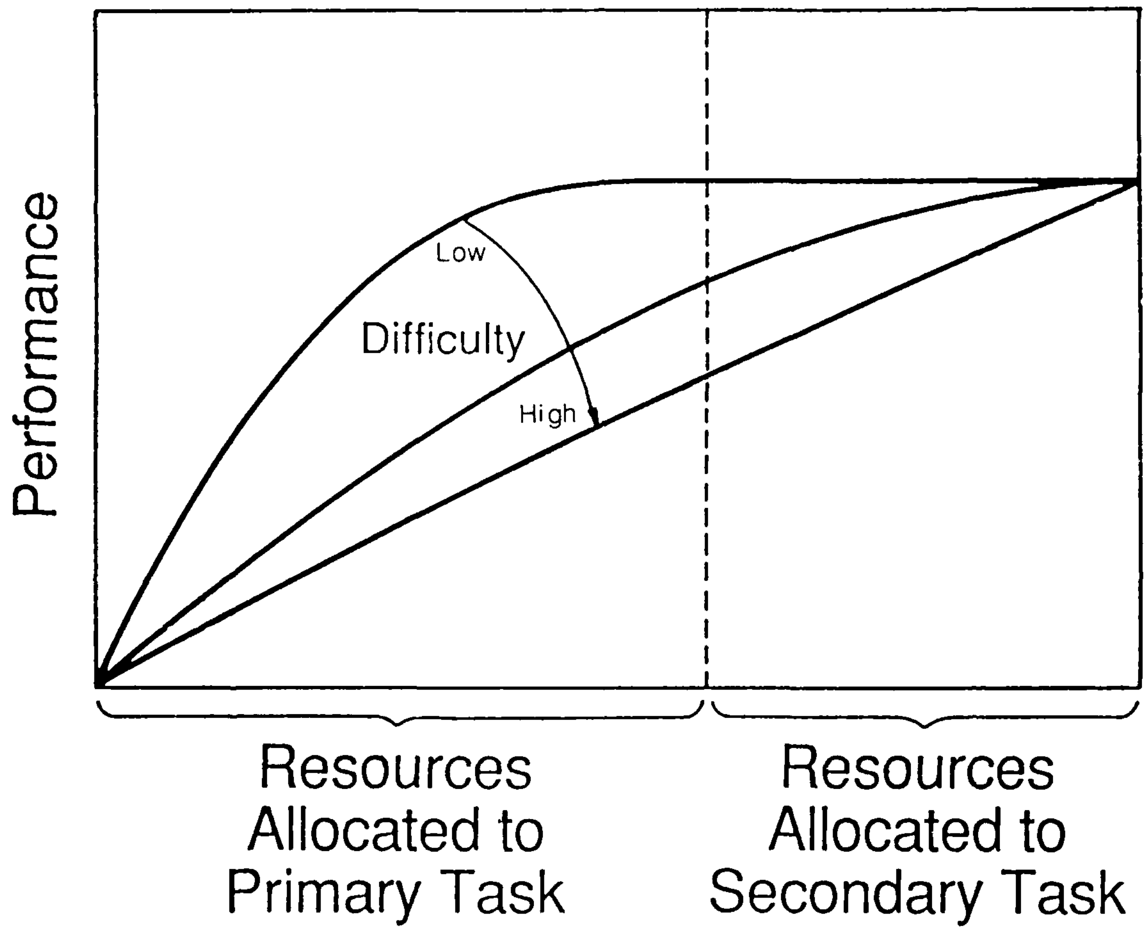


Figure 8.1. Graph of how performance is a function of the difficulty of primary and secondary tasks. (from Wickens, 1992)

curves. Note that there are no differences in maximum levels of performance between the intermediate and high skill level. But those with high skill will be able to perform more automatically, and will allow successful performance of concurrent tasks with the "residual resources." One important characteristic of human resources is that they exist in more than one variety. The specific nature of these "multiple resources" will be discussed in the following section on workload prediction.

Workload

Our discussion of attention and timesharing in the previous section has set the stage for the treatment of workload here. Figure 8.2 is one representation of

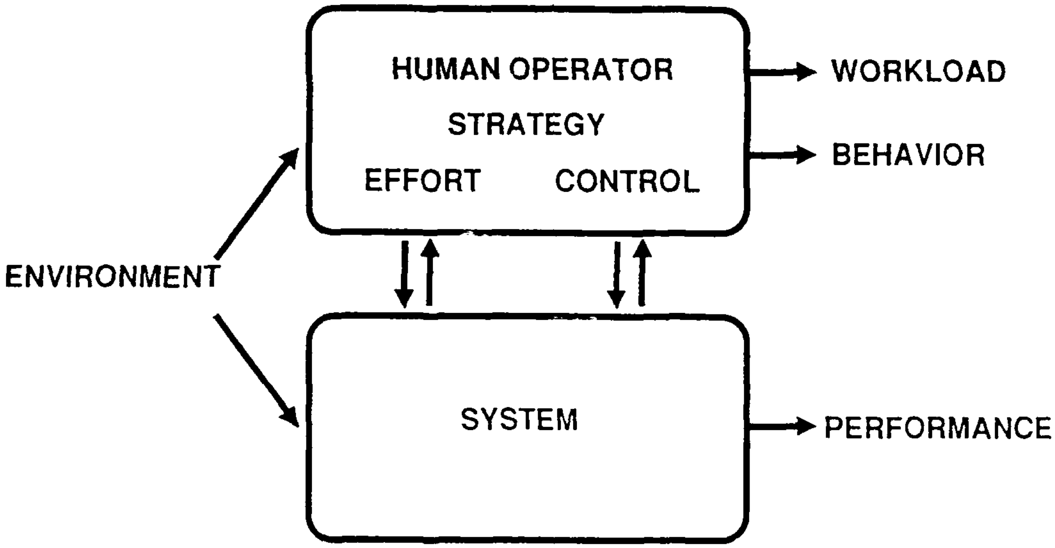


Figure 8.2 Model of workload.

workload. Loosely speaking, we can think of ***workload*** as the relationship between the capacity of a human operator and the demands of a system. That human operator interacts with the system in two ways. First, he or she is involved with control--doing things to it and watching what happens. Second, he or she is also involved with putting effort into this performance, and the system itself drains effort from the operator. The human and the system together work under the influence of an environment. The human outputs behavior. The system outputs performance. For example, in an aircraft, the human is doing things to the control yoke, and the aircraft is performing (i.e., following some flight profile). The human also outputs workload which is the experience of the effort involved in controlling or monitoring the system. This is what we measure when we measure workload, and these are the factors that basically drive workload.

There are a number of important case studies in which pilot workload has played a major role. Right now a major issue in the Army is whether one or two pilots should fly the LHX Light Attack Helicopter. That is very much of a

workload issue. Can one crew member manage the task load requirement with sufficiently low workload to make it fly satisfactorily with sufficient residual resources to handle the unexpected? An analogous choice was posed around 1980 regarding two- versus three-person flight crews on the generation of more automated commercial aircraft (e.g., the Boeing 757). The President established a workload task force to look at the issue of whether the flight engineer was necessary. The decision came down to allow two-crew operations, in part, because the mental workload was deemed to be allowable with this complement. FAR 25.23, Appendix D, talks about certifying aircraft for their workload. In such certification, workload estimations are used to compare systems. Does the old system impose less workload or more workload than the new system? Workload is also relevant in examining the impact of data-link based automation versus traditional communications with the air traffic control. Finally, there is the issue of using workload measures to examine the level of training of a pilot. As we saw in the previous section, although two pilots may fly the mission at the same level, if one flies with a lot less workload than the other, does that make a difference in predicting how the pilots will do later on or how well the pilots may transition from simulator training to the air?

What exactly is workload? How does workload relate to performance? How a plane performs in terms of its landing or deviation from the flight path tells you a good deal, but doesn't tell you all there is to know about the cost imposed on pilot workload by flying the aircraft. A good metaphor for workload is of a "dipstick to the brain." If workload depends upon this reservoir of resources we have, as shown in Figure 7.1, we would like to be able to push a little dipstick into the brain, find out how much workload there is, then just pull it out like we measure the amount of oil in a car. We'd like to be able to say the workload of this task is a 0.8 relative to some absolute capacity. This measure of ***absolute workload*** is a goal we are a long way from achieving. We will probably never be able to achieve it with a high degree of accuracy. Far more realistic is being able to make judgments of ***relative workload***; for example that the workload of the new system is less than or greater than the workload of the old system. This is different than saying the workload is excessive or not excessive.

In addition to the distinction between absolute and relative workload measures, a second distinction is between workload ***prediction*** and workload ***assessment***. A major objective of design is to be able to predict workload of an aircraft before flying a mission, as opposed to assessing the workload of the pilot actually flying. In this chapter we shall first contrast these two approaches: prediction and assessment. While our discussion in these sections will focus on conditions of ***overload*** (is workload excessive?), we will then turn to the other extreme of work ***underload***, and the closely allied issue of sleep disruption. Finally the

chapter concludes with a discussion of human error, a topic closely related to both underload and overload.

Workload Prediction

Timeline Analysis

The simplest model or technique for predicting workload is the ***timeline model.*** The timeline model is based on the assumption that during any flight task, the pilot, over time, performs a number of different tasks, and each task has some particular time duration. Therefore, we can estimate the workload on the pilot as being the proportion of total time that he or she has been occupied doing something. When applying this method, it doesn't matter what the difficulty of that task is. The only thing that matters is how long it takes to carry out the task. It doesn't make much of a difference whether two tasks are done at the same time or done at different periods of time. Timeline analysis has been developed extensively in the work that Parks and Boucek (1989) have done at Boeing, where they have developed specialized software for doing such analysis.

As shown in Figure 8.3, the ***Timeline Analysis Program*** (TLAP) simply codes a time record by lines, whose vertical position indicates the type of task, and whose length indicates the duration of time each task segment is performed. The time line is divided up into lengths of equal duration. Then the program sums within each unit of time the total amount of time the tasks are being done and the total time available. It computes the fraction of the time required to do each task and divides that by the time available within the interval. From that, the software comes up with a workload score for each interval.

The program can generate a chart for a particular activity that shows peaks and valleys. Figure 8.3 shows an example of a workload time history profile. Using such a technique, it is possible to establish a "red line" of absolute workload level, a workload you would say is "excessive." Then you can determine where design problems are in the epochs when the task demands exceed the red line. As one example, Parks and Boucek (1989) carried out an analysis of their view of the implication of the data-link system on flight crew workload. The scenario they fabricated was one with a weather deviation, an approach to landing, some major weather, a wind shear warning, missed approach, and a number of other events. They first traced out the pattern of activities carried out by the pilot-flying and the pilot-not-flying, under the conventional instrumentation and the conventional interaction with controllers. The task analysis was then repeated assuming their conception of the data-link system, which posited a data-link display on which, at the bottom of the CDU there was a message board that presented the necessary information from the data-link, (the automated

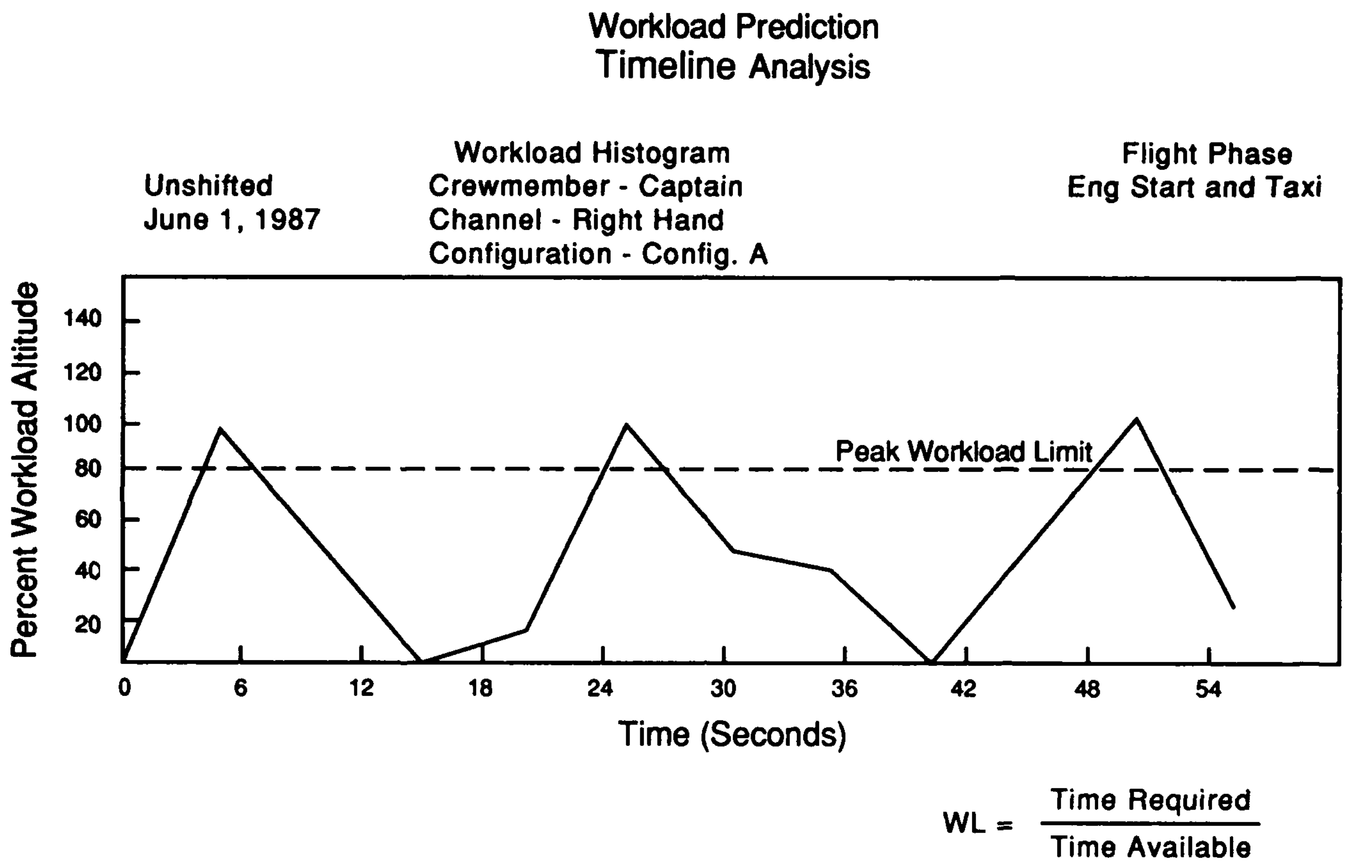


Figure 8.3. Example of workload time history profile as produced by Timeline Analysis Program. (from Parks & Boucek, 1989)

information given from the controllers).

The particular conclusions that they drew from this analysis are less important than the simple illustration of the technique. The way in which they applied it was one of looking at the change in workload for the copilot and for the pilot, from the conventional system to the data-link system. Using a more detailed analysis, they also broke down the tasks in terms of different channels of human resources that were loaded: internal vision, (vision that was head-down), external vision, the left-hand, the right-hand, cognitive activity, and "auditive activity" (listening and speaking). They found that with the data-link system for the copilot, there was a very substantial increase in internal vision; the eyes were much less frequently out the window and far more focused on head-down operations, because of the necessity of monitoring the CDU. Also, there was much more left-handed activity. There was also less auditory activity for the copilot, a reduction related of course to the decrease in voice

communications with ATC. A timeline for one of those particular channels, internal vision, is shown in Figure 8.4 for the advanced flight deck with a weather avoidance segment. Workload is plotted as a function of time in seconds. The heavy black line indicates an increase from the data-link system over the conventional system. The investigators found that at particular locations in time, something about the mission drove internal vision above the red overload line, where there is 100 percent workload (time occupancy). These events had to do with monitoring data-link for heading and altitude, concurrently with an instrument scan.

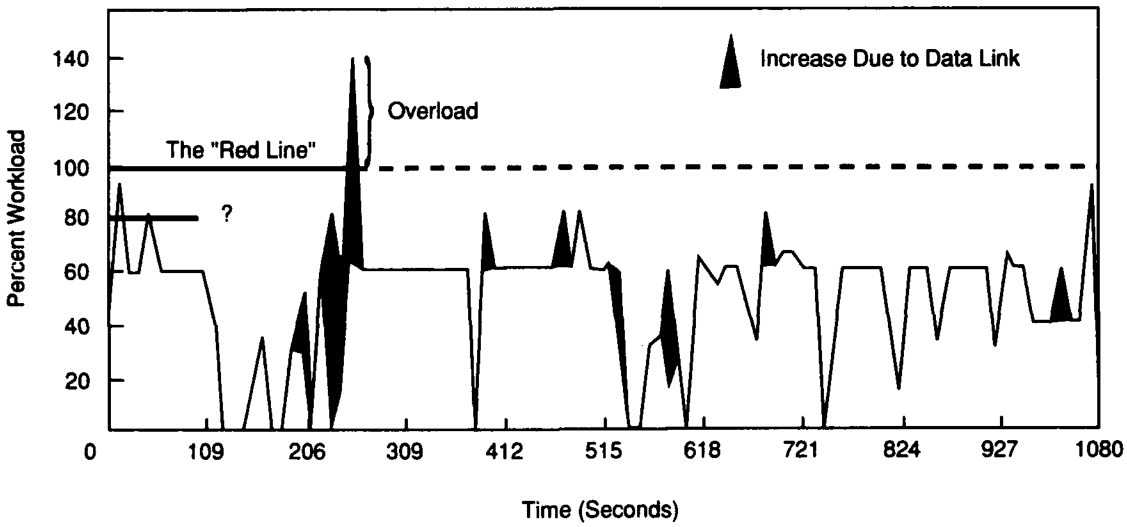


Figure 8.4. Pilot internal vision tasking in advanced flight deck for weather avoidance. (from Groce & Boucek, 1987)

There are some other examples of timeline analysis. For example, McDonnell Douglas has a slightly different version of a timeline program. Either version provides a good way of auditing what the tasks are and where the potential periods of peak overload may be. The technique has certain limitations however because it assumes that the workload of a task is only defined by how long it takes and not how intensive or demanding it is. We all know intuitively that there is a difference between how long something takes and how much demand it imposes on our mental process. For example, the pilot may have to retain three digits of information from ATC in short-term memory for five seconds, or

seven digits of information in short-term memory for five seconds. Either way, that task takes five seconds, but certainly keeping seven digits in mind is more demanding on our mental resources than keeping three digits in mind. Similarly, flight control with an easily controlled system may involve just as much stick activity but a lot less cognitive demand than flight control with a system that has long lags and is very difficult to predict. Timeline analysis doesn't really take into account the ***demand*** of the tasks.

A second problem is that the way timeline analysis is derived, the definition of a task is usually something you can see the operator doing, and it doesn't handle very well the sort of cognitive thinking activities that pilots go through (planning, problem solving), although timeline analysis is beginning to address this.

A third problem is that timeline analysis doesn't account for the fact that certain tasks can be timeshared more easily than others. Pilots can do a fairly good job of controlling the stick at the same time they are listening. Visual and vocal activity can be timeshared very easily. Visual and manual activities can be less easily shared. In other words, scanning the environment at the same time as entering information into a keyboard is much more difficult than speaking to a controller while looking outside the cockpit. Rehearsing digits is also quite difficult while talking or listening. Timeline analysis does not account for the fact that certain tasks are easy to timeshare and others are hard. These differences in timesharing will be elaborated below when we discuss multiple resources.

Finally, a fourth problem is that timeline analysis is fairly rigid. It sets up a timeline in advance and sees where different tasks will be performed, but in reality, pilots do a fairly good job of scheduling and moving tasks around. So if two tasks overlap in time according to the timeline set up by the analyst, pilots may simply postpone one in a way that avoids overlap.

Elaborations of Timeline Analysis

There are a number of more sophisticated workload prediction techniques that address some of these limitations of timeline analysis. Table 8.1 shows workload component scales for the UH-60A mission/task/workload analysis. It is an attempt by Aldrich, Szabo, & Bierbaum (1989), who have been working with the Army on the helicopter design to code the tasks in terms of how demanding or how difficult they are. The left column has a number for the difficulty scale of the task. A higher number means the task is more difficult. The first task on the list is "Visually Register/Detect (Detect Occurrence of Image)." It has a

difficulty value of 1. The authors have also defined six channels of task demand, analogous in some respects to the different channels used by Boeing.

Another way of accounting for the demands of a task is through a ***demand checklist***. That is, if you do an analysis of the task that a pilot has to do, there are certain characteristics of any given task that influence whether it is difficult or easy, independent of how long it takes. Consider, for example, the signal-to-noise ratio. It obviously is a lot easier to search for a runway if it is clearly defined than if it is partially masked by poor visibility. Other characteristics that influence display processing demand are the ***discriminability*** between different display symbols, the ***clutter*** on a display, the ***compatibility*** between a display and its meaning, as discussed in the earlier chapter, and the ***consistency*** of symbology across displays. Variables that influence the demand for central processing resources are the ***number of modes*** in which a system may operate, the requirements for ***prediction***, the need for ***mental rotation*** (as a pilot must often do when using an approach plate to plan a south-flying approach), the amount of ***working memory demands*** (time and number of chunks), and the need to follow ***unprompted procedures***. Demands on response processes are imposed by ***low S-R compatibility***, the ***absence of feedback from action***, and the ***need for precision*** of action.

Table 8.1
Workload Component Scales for the UH-60A Mission/Task/Workload Analysis

Scale Value	Descriptors
	Visual-Unaided (Naked Eye)
1.0	Visually Register/Detect (Detect Occurrence of Image)
3.7	Visually Discriminate (Detect Visual Differences)
4.0	Visually Inspect/Check (Discrete Inspection/Static Condition)
5.0	Visually Locate/Align (Selective Orientation)
5.4	Visually Track/Follow (Maintain Orientation)
5.9	Visually Read (Symbol)
7.0	Visually Scan/Search/Monitor (Continuous/Serial Inspection, Multiple Conditions)
	Visual-Aided (Night Vision Goggles [NVG])
1.0	Visually Register/Detect (Detect Occurrence of Image) With NVG
4.8	Visually Inspect/Check (Discrete Inspection/Static Condition (With NVG)
5.0	Visually Discriminate (Detect Visual Differences) With NVG
5.6	Visually Locate/Align (Selective Orientation) With NVG
6.4	Visually Track/Follow (Maintain Orientation) With NVG
7.0	Visually Scan/Search/Monitor (Continuous/Serial Inspection, Multiple Conditions) (With NVG)
	Auditory
1.0	Detect/Register Sound (Detect Occurrence of Sound)
2.0	Orient to Sound (General Orientation/Attention)
4.2	Orient to Sound (Selective Orientation/Attention)
4.3	Verify Auditory Feedback (Detect Occurrence of Anticipated Sound)
4.9	Interpret Semantic Content (Speech)
6.6	Discriminate Sound Characteristics (Detect Auditory Differences)
7.0	Interpret Sound Patterns (Pulse Rates, Etc.)
	Kinesthetic
1.0	Detect Discrete Activation of Switch (Toggle, Trigger, Button)
4.0	Detect Preset Position or Status of Object
4.8	Detect Discrete Adjustment of Switch (Discrete Rotary or Discrete Lever Position)
5.5	Detect Serial Movements (Keyboard Entries)
6.1	Detect Kinesthetic Cues Conflicting with Visual Cues
6.7	Detect Continuous Adjustment of Switches (Rotary Rheostat, Thumbwheel)
7.0	Detect Continuous Adjustment of Controls
	Cognitive
1.0	Automatic (Simple Association)
1.2	Alternative Selection
3.7	Sign/Signal Recognition
4.6	Evaluation/Judgment (Consider Single Aspect)
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (Consider Several Aspects)
7.0	Estimation, Calculation, Conversion
	Psychomotor
1.0	Speech
2.2	Discrete Actuation (Button, Toggle, Trigger)
2.6	Continuous Adjustive (Flight Control, Sensor Control)

Table 8.1 (cont'd)
Workload Component Scales for the UH-60A Mission/Task/Workload Analysis

4.6	Manipulative
5.8	Discrete Adjustive (Rotary, Vertical Thumbwheel, Lever Position)
6.5	Symbolic Production (Writing)
7.0	Serial Discrete Manipulation (Keyboard Entries)

(from Aldrich, Szabo, & Bierbaum 1989)

These are a series of guidelines that can be used to predict the amount of load on a task. There are other approaches to predicting task demand as well. Parks and Boucek have used an information complexity measure for computing task demands. However, what has been discussed up to now has still been a view of attention that really assumes that there is one pool of resources that are used for all tasks, or a series of separate and completely independent channels. That assumption of how the attentional system works is not in line with the fact that not all of the interference between tasks can be accounted for by difficulty. For example, entering data into a keyboard interferes a lot more with flying performance when it is done manually than when it is done by voice. When we change the structure of the task like this we can sometimes find a large difference in the amount of interference with flying. We also find another characteristic of dual task performance which indicates that not all tasks compete for the same resources, and this is called ***difficulty insensitivity.*** This is a situation when increasing the difficulty does not increase the interference with another task. Given the assumption that there is one pool of resources, then if we make one task more difficult, we pull resources away from the other task, and the performance of the other task ought to decline. But there are situations when this doesn't happen. For example, we can increase the difficulty of flying and a pilot's ability to communicate will not change much unless the flying becomes very, very difficult.

Multiple Resources

The above findings and others suggest that there is not a single pool of resources, but rather that there are multiple resources. So to the extent that two tasks share many common characteristics, and therefore common resources, the amount of interference between them will increase. For example, if we have two tasks that both demand the same resource, like controlling aircraft stability while adjusting a navigational instrument, there will be a trade-off in performance between them. However, if we have one task that demands resource A, and a second task that demands resource B, like listening, while flying a coordinated turn, there will be little or no mutual interference. As an analogy, if you have one home that relies on gas, and another home that relies on oil, there is not going to be any competition for heating resources between these homes if, say, the demand for gas suddenly increases.

A second characteristic of multiple resources is that we can talk about increasing the workload of a task, in terms of increasing the demands on a specific type of resource. If this resource is also shared with concurrent tasks, the difficulty increase will be more likely to lead to a loss of performance. In other words if two tasks demand the same resources, there will be a trade-off between the difficulty of one and performance of the other. If they use different resources, we can change the demand of one and not affect the performance of the other.

We have argued elsewhere that there are three distinctions that define resources. First, auditory resources are different from visual resources. Therefore, it is easier to divide attention between the eye and ear than between messages from two visual sources or two auditory sources. Second, the resources that are used in perceptual and cognitive processes in seeing, hearing, and understanding the world are different from the resources that are involved in responding, whether with the voice or with the hands. Third, we have contrasted spatial and verbal resources.

As we are perceiving words on a printed page or spoken words, we are using verbal resources. When employed in central processing, we use verbal resources for logical problem solving, rehearsal of digits or words, and mental arithmetic. For a pilot this could involve rehearsing navigational frequencies given by ATC or computing fuel problems. Anything that has to do with the voice uses verbal response resources.

In perceiving spatial information, we do a variety of things. We do visual search; we process analog quantities like moving tapes or moving meter displays. We also process flow fields, that is, estimate the velocity over the ground, from the flow of texture past the aircraft. We recognize spatial patterns on maps, to help form a guidance of where to fly. Spatial central processing involves imagining the airspace, or mentally rotating maps from say a north-up to a heading-up orientation. Spatial responses are anything that involves manually guiding the hands, fingers, feet or eyes through space: using the control yoke, the rudder pedals, and the keyboards or engaging in visual search.

Thus the idea behind multiple resources models is that you can predict how tasks will interfere with each other or how much workload will be experienced not only by how long those tasks take to perform and by how demanding those tasks are, but also by the extent to which two tasks demand common resources. There are now a number of different efforts in the research design community, more directly focused on military systems, that have elaborated upon versions of multiple resources theories to come up with computation models that will take a timeline and a task demand coding, and make predictions of the workload on

the pilot. Both Honeywell and the Boeing people have been involved in developing a model of this sort (North & Riley, 1989).

Workload Assessment

A framework for understanding workload assessment is presented in Figure 8.5 which shows a graph that presents across the bottom line the resources demanded by a task or set of tasks. The farther to the right on this axis, the more the pilot is having to do more tasks or perform tasks that are more difficult. The pilot has available multiple resources that can be given to those

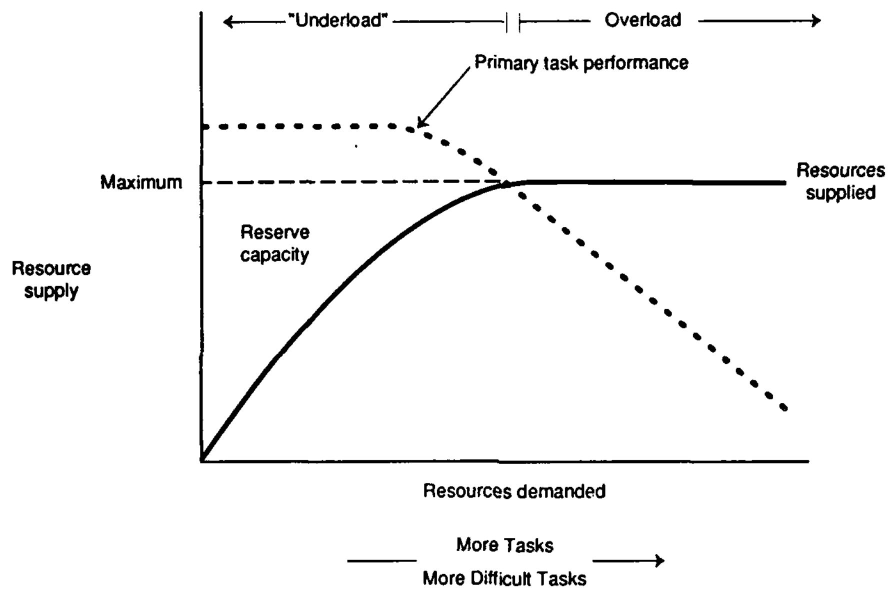


Figure 8.5. Graph showing workload assessment. (from Wickens, 1992)

tasks. These can be supplied up to a maximum, shown as the horizontal dashed line. As the graph moves from doing nothing at all (on the left end) to doing something that is moderately difficult at the middle of the graph, more resources are demanded but the pilot can adequately supply those resources, so there is a nice linear supply-demand curve. As long as this linear function remains, resource supply keeps up with demand, and the pilot's performance is going to be perfect. This region where supply satisfies demand is called the "underload" region. By underload, we don't really mean the region of boredom where the pilot is doing nothing at all, but rather the region where he is not asked to do more than can possibly be done.

If you look at all how well the pilot is performing the task at hand when demands are in the left side of the graph (e.g., maintaining the flight path), what you will see is less and less reserve resources available to do other things. As we push the demand beyond the maximum supply at the middle of the figure, the pilot is getting into the "overload" region. There is an excess of demands and the pilot needs more than he can give. As a result, performance of the task of interest is going to begin to deteriorate. The measurement of workload requires looking across this whole range of task demands, from underload to overload. This suggests that how we measure workload may vary depending on where the pilot falls in the underload and overload regions. At the left, we must measure ***residual resources***. At the right, we may measure performance directly. Four major techniques of measuring workload are generally proposed: measuring the primary task itself, measuring performance on a secondary task, taking subjective measurements, and recording physiological measurements.

Primary Task Performance Measures

In aviation, the critical primary task is flight performance. How well is a pilot actually doing keeping the plane in the air along a predefined flight path trajectory? The direct measure of primary task performance might be some measure of error or deviations off of that trajectory. However, it is also important to measure not only performance, but some index of control activity; that is, how much effort the pilot is putting into keeping the plane on the trajectory. We need to measure control activity because we can get two aircraft that fly the same profile with the same error, but one requires a lot of control activity and one needs very little control activity. It turns out that one good measure of control activity is the ***open loop gain***, which is the ratio of the pilot's control output (yoke displacement) to a given flight path deviation.

Figure 8.6 shows the relationship between gain (effort) and error. The upper left box represents a timeline of a pilot flying a particular profile under low workload because there is little error and little control effort being made. This is an unambiguous measure of low workload; performance (flight path error) is good and effort is low. In the upper right box, we have a situation where the error is low but the pilot is putting in a lot of control activity to maintain that low error. We would see there is a high gain or high effort invested in the flight performance. This is probably a high workload situation and suggests that there is some sort of control problem. That is, some sort of problem in the way the information is represented or the handling of the aircraft, so it is taking a lot of effort to keep the plane flying steadily. This situation may also reflect flying in high turbulence.

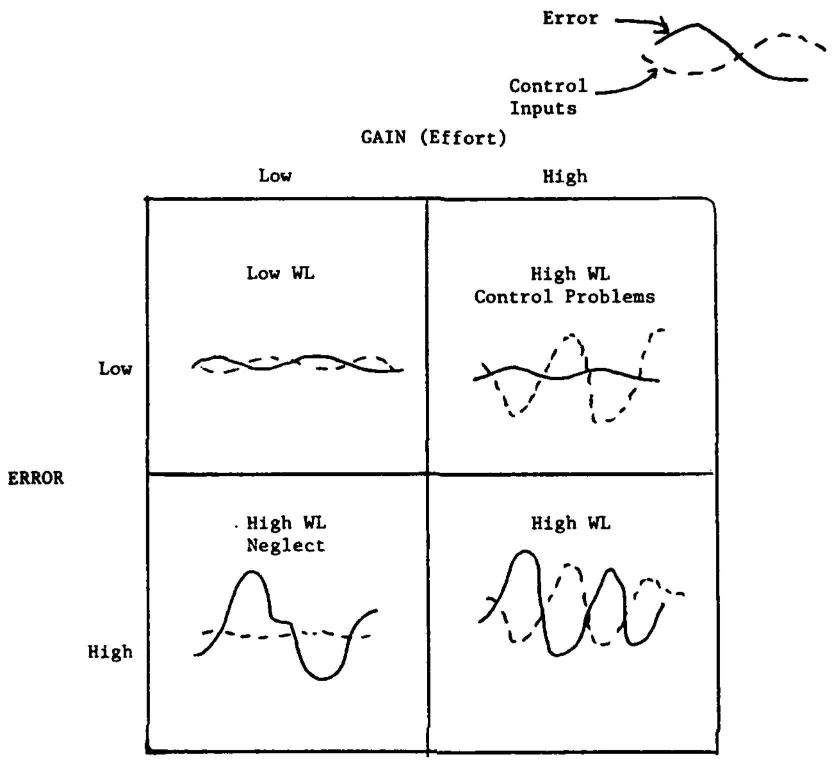


Figure 8.6. Relationship between gain and error. (original figure)

In the lower left box is represented the opposite situation in which there is not much control activity going on, but there is a fairly high amount of error. It is almost as if the plane is flying through turbulence and the pilot is not doing anything with the stick. This pattern may very well signal neglect, where the pilot is neglecting the flight control and allocating resources to something else--system problems or problems with other aspects of the aircraft. It is also an indicator that there is high workload, but the high workload is not associated with the flight control itself, but with some aspect of the aircraft environment. Finally, the lower right box shows the worst situation, in which the pilot is producing a lot of control activity and is still generating a lot of error for whatever reason. Thus there is very high workload in this situation.

The important point illustrated in this figure is that looking at performance of the primary task itself as an indicator of workload is not sufficient. You have to look jointly at performance of the system and at the behavior of the pilot.

Secondary Task Performance

A second approach to workload measurement is the secondary task. This technique assesses the extent to which the pilot has enough residual resources to perform another, secondary task at the same time as a primary task without letting performance in the primary task drop. When doing a difficult primary task, if we give the pilot a secondary task, he is going to either have no resources for that secondary task, or, if resources are diverted, the primary task is going to drop (Wickens, 1991).

One example of a secondary task is ***time estimation***. Suppose the pilot is flying along and is asked to give a voice report every time he thinks 10 seconds has passed. Time estimation generally becomes more variable and the intervals longer as the workload increases. Another secondary task that has received a fair amount of interest is the task of a ***memory comparison***. While flying along, the pilot hears a series of probe signals. Maybe they represent call signs. Every time he hears the call sign of his own aircraft, he presses a button. Every time he hears the call sign of another aircraft, he does nothing. So he compares each call sign to his memory. If it matches he responds. This task is sometimes called the ***Sternberg Task***. The response time to acknowledge call signs is longer with higher levels of workloads. ***Random number generation*** is another possible secondary task. The pilot is asked to generate a series of random numbers and the more difficult the primary task, the less random the numbers become. Another secondary task is the ***critical instability tracking task***, in which a second tracking task is built into the pilot's primary flight control loop. Error on this task directly reflects the difficulty of the flight dynamics of the primary task.

All of these types of secondary tasks have various problems. One problem they have in common is that they are all sensitive to multiple resources. If you have a secondary task that demands resources that are different from the primary task, you are going to underestimate workload. If you have a primary task that is heavy, in terms of perceptual-cognitive load--rehearsing digits would be a good example--and you have a secondary task that is heavily motor, like performing a critical tracking task, it is like you are looking in one corner of a room for something that exists in a different part of the room. So you need to have your secondary tasks demand the same resources as the primary task.

Perhaps even more critical, at least for in-flight secondary task measures of mental workload, is this problem of ***intrusiveness***. We can all imagine the resistance that a pilot would give if he were trying to fly the aircraft through high workload conditions, and at the same time had to generate a continuous stream of random numbers, or had to continuously control a side-tracking task. He simply wouldn't want to do it. This is the biggest bottleneck towards the

introduction and the use of secondary tasks--they tend to be intrusive into the primary task and disrupt the primary task; and this is a major problem when the primary task is one involving a high-risk environment (i.e., in flight recording, rather than simulation).

A solution to the problem of intrusiveness is a technique called the ***embedded secondary task***; that is, use of a secondary task which is an officially designated part of the pilot's primary responsibilities, but is fairly low in the hierarchy of importance for the pilot. In flying, there is a certain intrinsic task priority hierarchy. For example, there is the standard command hierarchy to aviate, navigate, and communicate in that order of priority. With more precision we can further rank order tasks in terms of those that have very high priority, say maintaining stability of the aircraft, those of extremely low priority, like answering service calls from the back of the aircraft, and those things in between. The idea behind this prioritization scheme is that as the workload increases from low to high, the lowest priority tasks are going to drop out, so when the workload is very, very high, the only thing that will be left to do is the highest priority task. Thus good embedded measures of secondary tasks are those tasks that are naturally done but are lower down in the priority hierarchy. An example might be acknowledging call signs. To the extent that this is a legitimate part of the communication channel, one can measure how long it takes the pilot to acknowledge the call sign as an embedded secondary task. Our research has indicated that airspeed control is a good embedded secondary task. The control of airspeed around some target is of lower priority, or at least seems to be reduced in its accuracy more, when the demands for the control of the innerloop flight path error, (heading and altitude error), become excessively difficult. So as the demand goes up, the airspeed errors seem to increase, more so than do the other types of errors.

Subjective Measures of Workload

The third category of workload measures, which is often the most satisfactory to the pilot, is the subjective measure. There are a number of different techniques of subjective workload measurement. One is a unidimensional scale. An example of this is the ***Bedford Scale*** shown in Figure 8.7a, and involves a decision tree logic. There are a series of questions: Was workload satisfactory without reduction? Was workload tolerable for the task? Was it possible to complete the task? If the answer is yes or no, then you go on up to some higher levels that eventually allow you to categorize the workload of a task on a 10-point scale. Similar to the Bedford Scale is the modified ***Cooper-Harper Scale*** (Figure 8.7b), which is taken more directly from the Cooper-Harper scale of flight handling quality, but now has questions phrased in terms of workload. The important point is that you can get a single number, and that number is guided by a

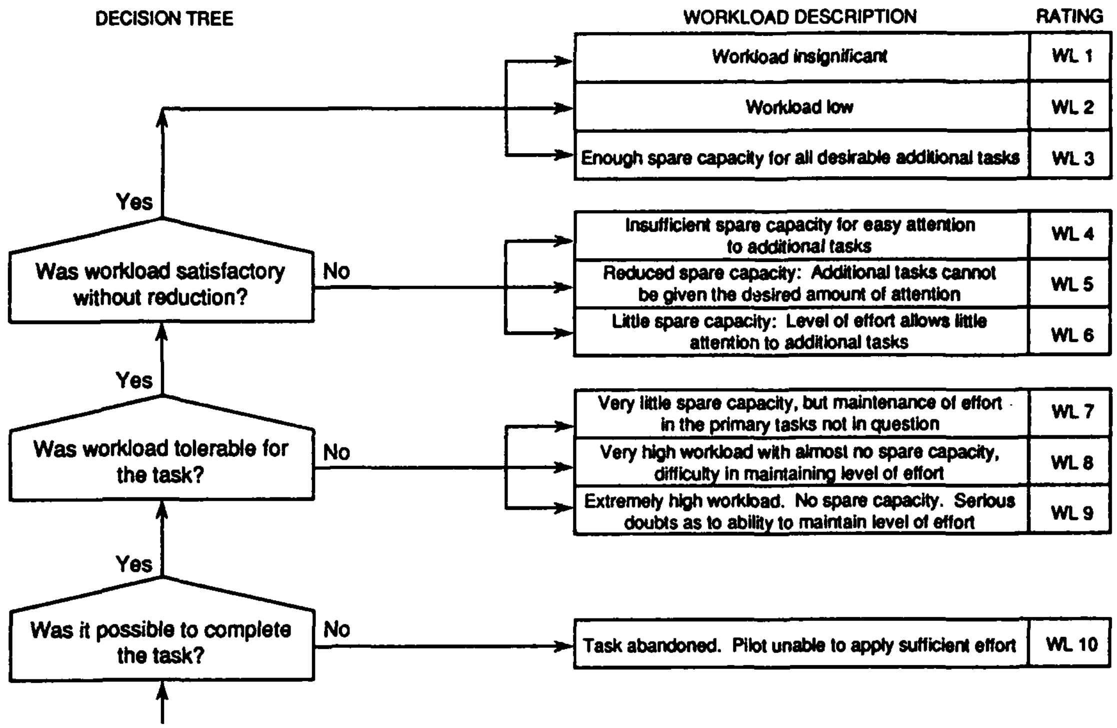


Figure 8.7a. The Bedford pilot workload rating scale. (from Roscoe, 1987)

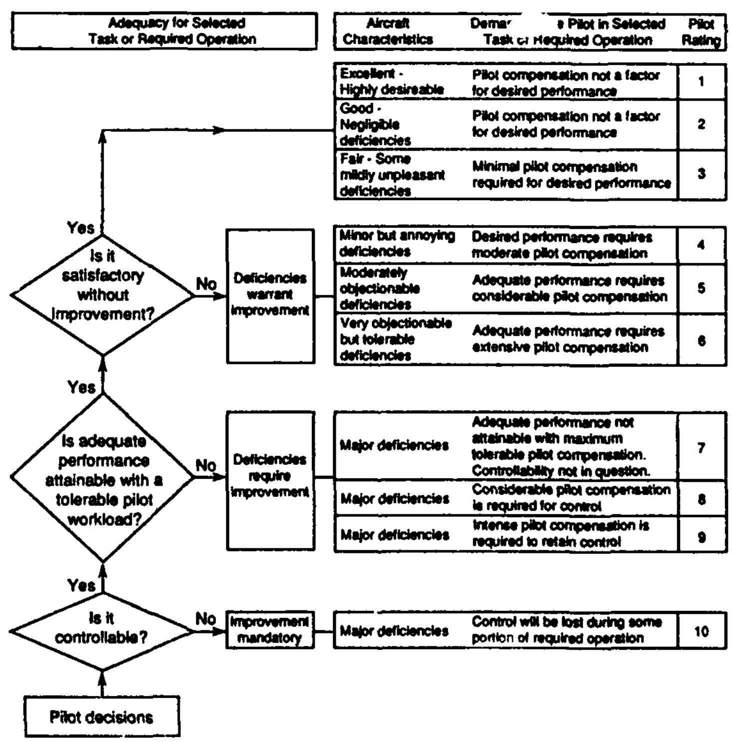


Figure 8.7b. The Cooper-Harper pilot workload rating scale. (from Cooper & Harper, 1969)

series of verbal decision rules about how it is that you ought to interact with the task. Both of these unidimensional scales, the Bedford and the modified Cooper-Harper, are simple. Because they are simple, they have a certain amount of ambiguity. It is not always clear why a task is rated difficult, because the scale won't tell you if it is difficult, for example, because it had difficult response characteristics, or because the displays were hard to interpret, or there was heavy time pressure or heavy cognitive demands, etc.

Multidimensional scales, in contrast, assume that there are several dimensions underlying subjective workload, and reveal what these dimensions are. The two major candidates for multidimensional scales are the **Subjective Workload *Assessment Technique*** (SWAT) and the ***NASA TLX Scale***. The SWAT, which was developed for the Air Force at Wright Patterson AFB, assumes that we experience workload in terms of three dimensions: the time demands of the task, the effort of the task, and the stress the task imposes on us. It asks the pilot to indicate for each of these scales, on a three-point rating, whether the time, effort, and stress levels are low, medium, or high. By a fairly elaborate procedure which uses all 27 possible workload ratings derived from low, medium, and high combinations for each of these three scales, it is possible to determine which scale is more important for a particular pilot. This procedure is used as a way of coming up with a single measure of workload from these three ratings on each of the different scales.

Two major problems have been found with the SWAT technique. The sorting procedure it uses, which seems to be a mandatory part of SWAT, is time-consuming. The other problem has to do with the scale resolution; that is, SWAT only allows you to say that workload is low, medium, or high on each scale. If you consider your own flight experience, you are able to give a lot more precision to workload than three levels. You have more power of discrimination between the resource demands of the task than simply low, medium, and high. What happens when only three rating levels are available is that people tend to choose the middle level, and pretty soon you don't get much resolution at all.

A different technique, as an alternative to the SWAT is the NASA Task Load Index, or TLX scale. This was developed by Sandra Hart at NASA and assumes that there really are six dimensions of subjective workload: mental demand, physical demand, temporal (time) demand, the level of performance the pilot thinks he or she has achieved, amount of effort, and frustration level with the task. For each of these, there is a verbal description of what it means, and, furthermore each of these different demand levels can be rated on a 13-point scale. You do it by putting a mark on a piece of paper somewhere along the 13-point scale. The scale gives the pilots more freedom and flexibility to rate on

different dimensions without a lot of extra effort, and probably provides more information. In fact, some comparisons of how well the two different scales have differentiated loads indicates that the TLX scale does a better job than the SWAT. TLX also has a procedure that allows the six dimensions to be combined into a single workload rating. For many purposes, the single-dimensional rating scales are probably adequate for picking up most of what there is in workload.

There are really three problems with subjective workload measures. One of them is ***response bias***. If you are simply asking for a rating of workload, we all know there are individual differences among pilots. One may not ever admit that the workload is greater than three, no matter how difficult things are. Another may be very quick to admit to high levels of workload whether they exist or not. A second problem with subjective workload measures is related to memory. An example would be if we were evaluating two tasks, flown on two different systems, and the pilot is asked to compare their workload. Since the pilot's memory for the first one may have degraded, he may not be able to make an accurate judgment based on memory. The third problem with subjective workload measures is that they do not always agree with performance. It sometimes happens that when two systems are compared, one gives better performance than the other. However, the one that gives better performance is, in fact, shown to have higher measures of subjective workload. Which measure should then be trusted by the designer?

Physiological Measures of Workload

The fourth category of workload measures are physiological measures. Several of these have been proposed: heart rate (both mean rate and variability), visual scanning, blinking and various measures of electroencephalogram (EEG) that can measure fatigue and, finally, the evoked potential, the momentary changes in the EEG that are caused by a discrete event, like the sudden onset of a light or a tone. The prevailing view is that most of these techniques have some uses, but as far as being reliable measures of pilot workload, particularly in civil aviation, there are more problems than there are benefits. The most successful measures appear to be those that relate to heart rate. Here, there are two specific measures. There is the mean heart rate. That is, the number of beats per minute. The faster the heart beat, presumably the higher the level of mental workload. That does hold true more or less, but there are other factors, unrelated to mental workload that cause the heart to beat fast. Certainly two of these are arousal and stress. Another one is simply physical load. So in a physically taxing environment, even though the mental workload may be low, the heart rate may still be very rapid. Thus the mean heart rate is not a terribly good indicator of mental workload by itself.

A better measure of cognitive load is the variability of the heart beat interval (Vicente, et al. 1987). It has been found that as the workload gets higher, the variable of the heart gets lower. Figure 8.8 shows some data taken at Wright Patterson (Wilson, et al. 1988). It is a timeline of two minutes which plots at the bottom, the interval between each heartbeat. The fact that the curve oscillates suggests that the heartbeat interval is itself variable. Some periods

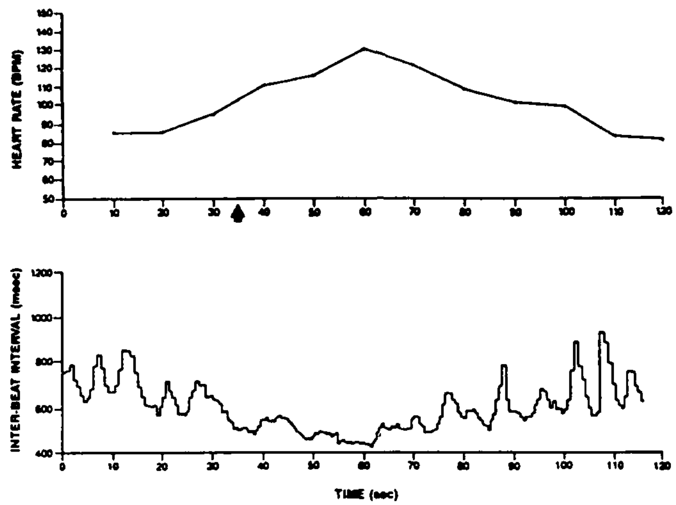


Figure 8.8. **Graph plotting inter-beat time intervals for heartbeat over a two-minute period. A birdstrike appears at approximately 35 seconds. (from Wilson, et al. 1989)**

the beats are close together, then they get slower, then they get faster, then they get slower. So this oscillation represents variability in the inter-beat interval. The overall level represents the overall inter-beat interval or the mean heart rate, plotted at the top. When the level is low, that means the heart is beating very fast. In the figure, note that at 35 seconds into the flight test, a bird struck the windshield. This was a fairly traumatic event, and you can see very dramatically an increase in heart rate (decrease in the inter-beat interval) and a reduction in the variability. So both emotional stress and the cognitive load of dealing with this unexpected event made the heartbeat faster and caused much less variation. Figure 8.9 (top) shows another case of relatively low variability in heartbeat, indicating high workload. Figure 8.9 (bottom) shows the change from high to low variability (low to high workload) with little corresponding change in emotional load.

Collectively, it is hard to say which technique of workload measure is best. In civil aviation tests by both AirBus Industries and Douglas there has been some success with the physiological measures. The best approach is probably one that

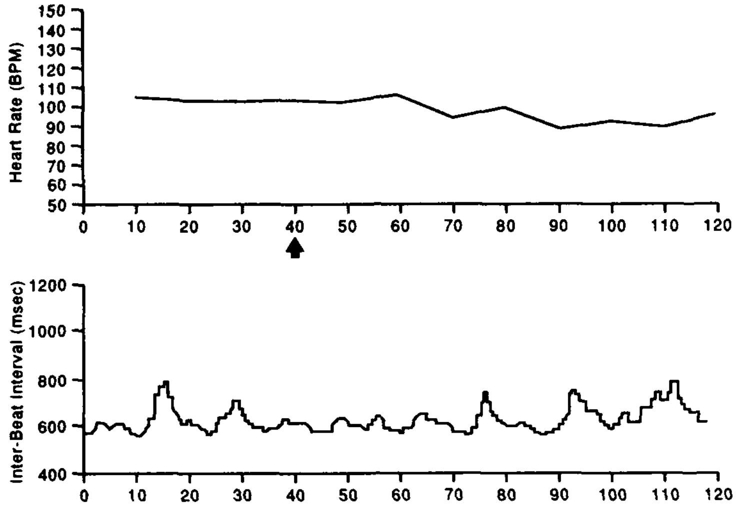


Figure 8.9. Graph plotting inter-beat time intervals for heartbeat over a two-minute period. Note the reduction in variability at t=40, with no corresponding change in mean heart rate. (from Wilson, et al. 1988).

involves comparisons across primary task performance measures, and embedded secondary tasks, augmented by subjective and possibly physiological measures, with an emphasis on the heart rate measures.

A Closed-Loop Model of Workload

The traditional view of workload has involved a fairly static concept expressed in Figure 8.10a, which proposes that there are a certain number of things that we could call drivers of workload. These are things that vary in a task or environment to increase the workload. Drivers of workload are task requirements, available resources, time available, and operator experiences. Drivers imposed on a task change the physical and mental actions required for the task and produce workload and performance as a result. This is an ***open-loop approach*** to workload. Simply stated, something is done to the operator, and it produces workload.

More recently, a dynamic ***closed-loop*** concept of workload has been proposed (Hart, 1989). This is illustrated in Figure 8.10b. The FAA, NASA, and the Air Force have cooperatively sponsored a program to look at workload as a more dynamic and adaptive phenomenon. As in the static concept, all of the drivers of workload are again represented. But there are also a set of fairly sophisticated cognitive activities, assumed to be carried out by the pilot. These include planning, setting priorities, establishing a schedule, allocating effort, focusing attention on certain tasks, ignoring others, etc. As a result of this adjustment, the pilot experiences some mental and physical demands, which we call workload, but the workload experienced at one moment in time is used to continuously adjust performance, establish priorities, and change task

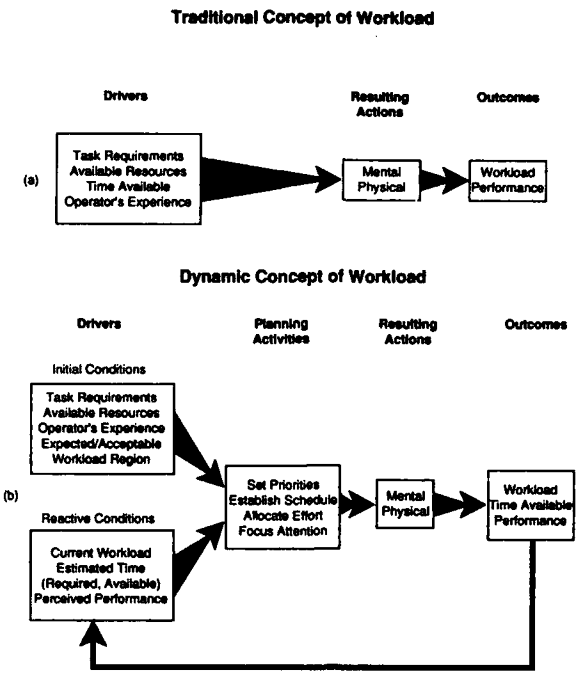


Figure 8.10. (a) Static and (b) dynamic concept of workload. (from Hart, 1989)

scheduling. Stating it in another way, people don't just experience workload then express it. Instead, if they experience workload, and the workload is too high, they drop tasks. If the workload is too low, they assume tasks.

Unfortunately, we really do not have a very strong database on how well people conform to this model. For example, there aren't good data regarding how good a job people do at shedding tasks appropriately and knowing whether optimal task shedding is done well under normal conditions, or done poorly under stress. A program of research at NASA and the Air Force is beginning to examine this issue, and there is a similar research program at Illinois to investigate task shedding.

One important implication of the closed-loop model, which we have not yet addressed, is that as people become underloaded they will tend to assume "pick up" tasks. The goal of a pilot is not to minimize workload, but rather to keep workload at some moderate, stable, intermediate level. This obviously has long-term implications for the system designer who is considering the appropriate level of automation. The goal of automation should not be to eliminate the pilot and reduce the pilot's workload to zero, but rather to simply address the overload conditions, and consider problems of the underload conditions as well. There has been a slight disconnect between the approach that more automation is invariably better, and the approach that automation ought to be designed to keep workload at an intermediate level rather than to eliminate all tasks from

the pilot's repertoire. The problems of excessively low workload, and their close relation to issues of sleep disruption will now be addressed.

Underload

The flip side of high workload is ***underload***. As we discuss underload in this section, it refers to situations of long periods of relative inactivity. Transoceanic flights or long cross-continental flights are examples of underload, where very little is actually happening. It is not surprising that very long periods of low workload really are not optimal. The pilot will try to create some level of workload, whether it is flight-related or not, in order to avoid sleeping. Some interesting studies of air traffic controllers by Paul Stager in Canada found that a predominance of ATC error seems to occur at relatively low workloads rather than periods of high overload.

One of the things we know about low workload periods is that these interact negatively with sleep loss. Pilots under sleep loss conditions are much more likely to perform poorly under low workload periods than pilots who are well rested, and so we now turn to a discussion of this important topic.

Sleep Disruption

There have not been many systematic studies of the effects of sleep deprivation on pilots' performance. Perhaps the best of these was a study carried out by Farmer and Green (1985) in the UK, in which they worked with 16 pilots. The pilots were deprived of one night's sleep, by being kept awake for 24 straight hours. Then they did a series of in-flight maneuvers, with a wide-awake check pilot to make sure that nothing disastrous happened. Farmer and Green looked at the kind of errors that were made, and found that the errors occurred mostly during the low activity portion of the flight, at the times when not much was going on, except for an occasional need to respond to, for example, unpredictable and infrequent warning signals. These are what psychologists call the "vigilance tasks."

Characteristics of Sleep

Because we know that sleep loss has consequences that are harmful in low workload environments, it is important to understand some of the characteristics of sleep. We have two different forms of sleep. One is ***rapid eye movement*** (REM) sleep in which the eyes are twitching, there is a lot of dreaming, and there is actually a fairly high level of brain activity. The other is ***slow wave sleep***, so named because the EEG is very slowly changing during this type of sleep. The brain is very quiescent during slow wave sleep. There is not much

dreaming activity going on. REM sleep takes place later in the night. Slow wave sleep takes place predominately during the first part of the night. There is good evidence that both kinds of sleep are important for the overall health of the individual.

The whole sleep wake cycle is defined not only in terms of staying awake and being asleep, but also by a set of body rhythms, called ***circadian rhythms*** that reflect different characteristics of the efficiency of performance. These circadian rhythms run on a 24-hour cycle and can be defined by body temperature, the depth of sleep, sleep latency, and performance. Figure 8.11 shows the average duration of sleep episodes and the body temperature of a person during a 48-hour time period.

What the function shows is that temperature is lowest in the night and the very early morning period. It begins to climb during the day, reaches its peak in the late afternoon and evening, then declines at night. The graph of temperature coincides with the bar graph that plots the duration of sleep. This graph shows

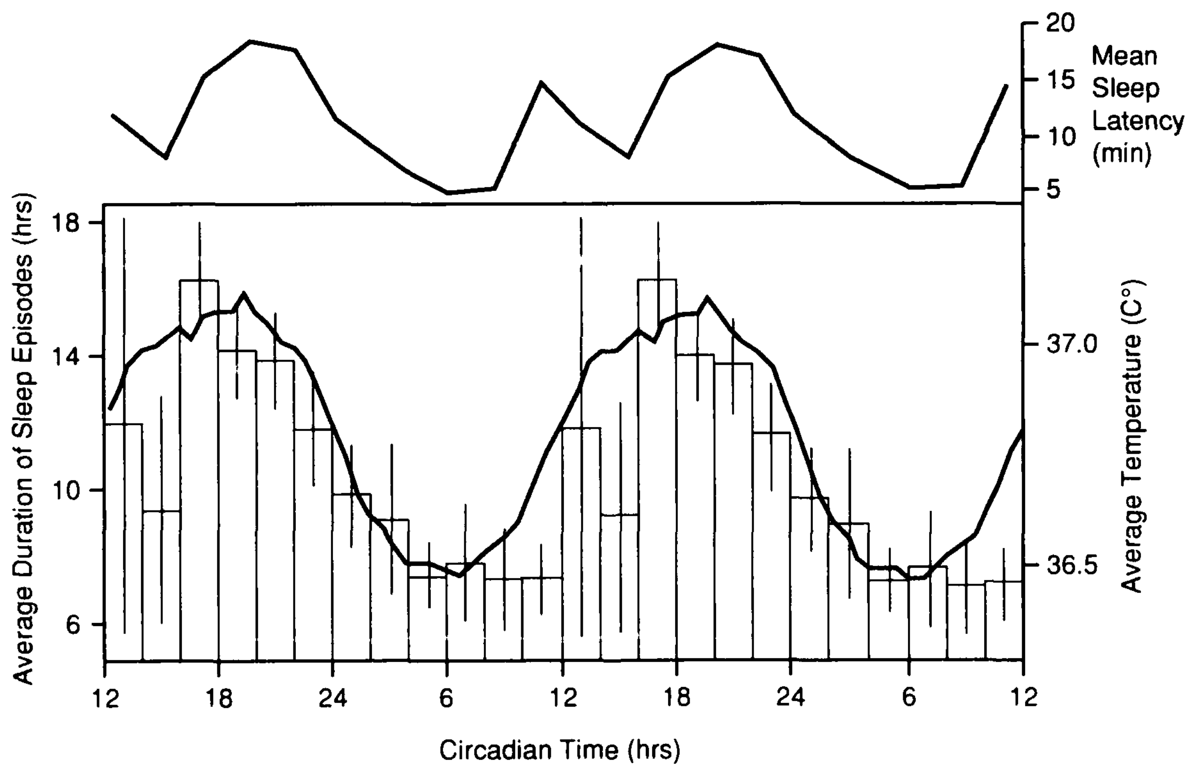


Figure 8.11. Graph of sleep duration and its relationship to circadian rhythm. (from Czeisler et al., 1980)

that if you go to sleep sometime in the early morning hours, your sleep duration will be relatively short. If you go to sleep during the evening, your duration of sleep will be longer.

A third characteristic of the circadian rhythms has to do with ***sleep latency***. Figure 8.12 shows a graph of the mean sleep latency of subjects who received the Sleep Latency Test. Sleep latency is how long it takes you to fall asleep. If there is a long latency, it means you are wide awake, and so you are not about to nod off to sleep. If there is a relatively short latency, it means you are very prone to fall into a deep sleep. Figure 8.12 covers results of a 24-hour period from 9:30 am to 9:30 am. Eight 21-year-old subjects and eight 70-year-old

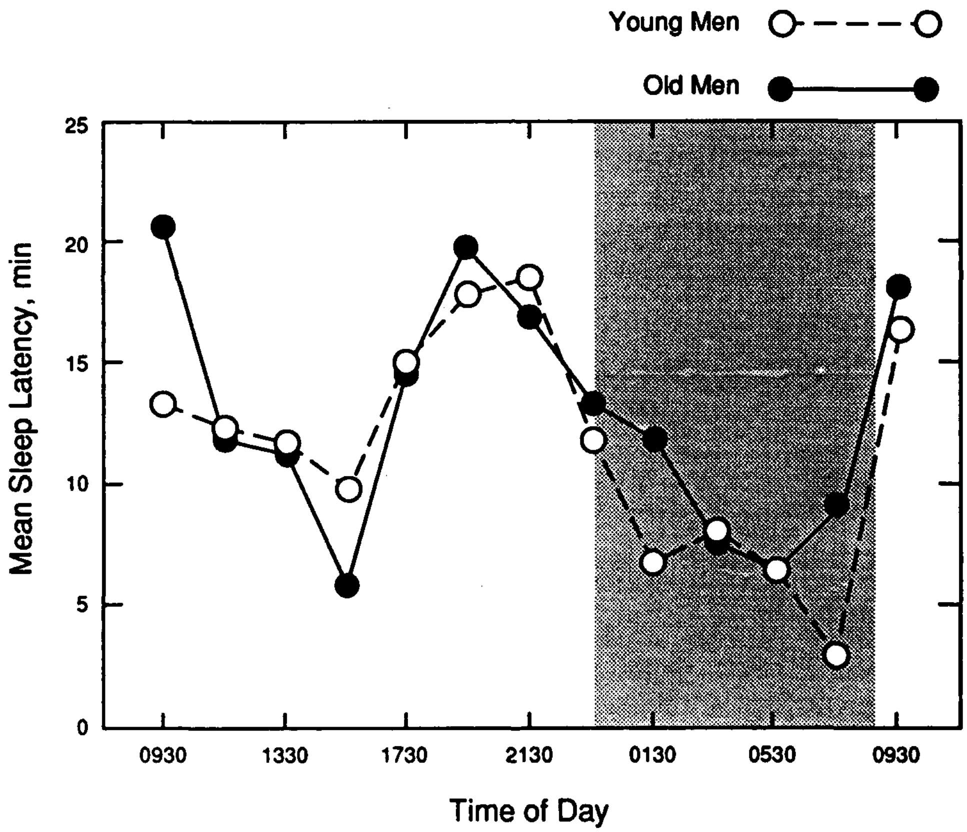


Figure 8.12 Mean sleep latencies for 21-year-olds and 70-year-olds. (from Richardson et al., 1982)

subjects received the Mean Sleep Latency Test (MSLT), while awake, during the day, followed by four brief awakenings at 2-hour intervals during the night (shaded). In the afternoon, there is a "post-lunch dip" which indicates that in the afternoon we tend to fall asleep and drop off rapidly. Sleep latency gets longer in the evening time (it takes longer to fall asleep), but then again becomes very short in the morning, and rises again during the daytime. The measures of temperature and sleep duration show only one cycle during the day, while sleep latency has the same general cycle but with this little extra dip in it in the afternoon.

Performance is the all-important measure related to sleep deprivation. Figure 8.13 shows how human performance on various tasks changes during the day. The performance tends to correspond with body temperature, but also shows hints of the "post lunch dip" characteristic of sleep latency. One graph shows psychomotor performance, like a tracking task. You do progressively better during the day, best in the early afternoon, and do relatively poorly at night and in the early morning hours. The other graphs show the measurement of reaction time, and of ability to do symbol cancellation and digit summation. The collective implications revealed by all of these effects is that we have a regularly trained rhythm that describes how fast we go to sleep, how long we sleep, our body temperature, and the level of performance, all of which show a very pronounced dip in the time from midnight until about six in the morning. The data strongly suggest that when possible, flight schedules ought to be arranged to take advantage of the capacity for sleep. Flight schedules that allow pilots to sleep at times when they go to sleep fastest and sleep for the longest are better than those that give pilots the opportunity to sleep at times when they have a hard time sleeping because their sleep latency is long.

Sleep Disruption in Pilots

A lot of the research on sleep disruption has either been based upon subjects that were not pilots, or were military pilots, so there are not a lot of data that generalize directly to civil aviation. There are two important studies that were carried out at NASA that do have a direct bearing on the civilian piloting community (Graeber 1988). One of these is a short-haul study in which a large number of pilots were evaluated during a series of domestic short hauls. They flew for three or four days before returning to the home base. Out of that study came the first systematic conclusions of the effects of sleep cycle on the short haul. First, the pilots began the trip with a sleep loss, because they were apt to sleep less than the normal amount the night before they took off for the first leg. Thus they started out behind the eight ball. This is interesting, because it is precisely the opposite of a concept that has proved to be an effective antidote against sleep loss, the concept of ***prophylactic sleep***. This is defined as getting

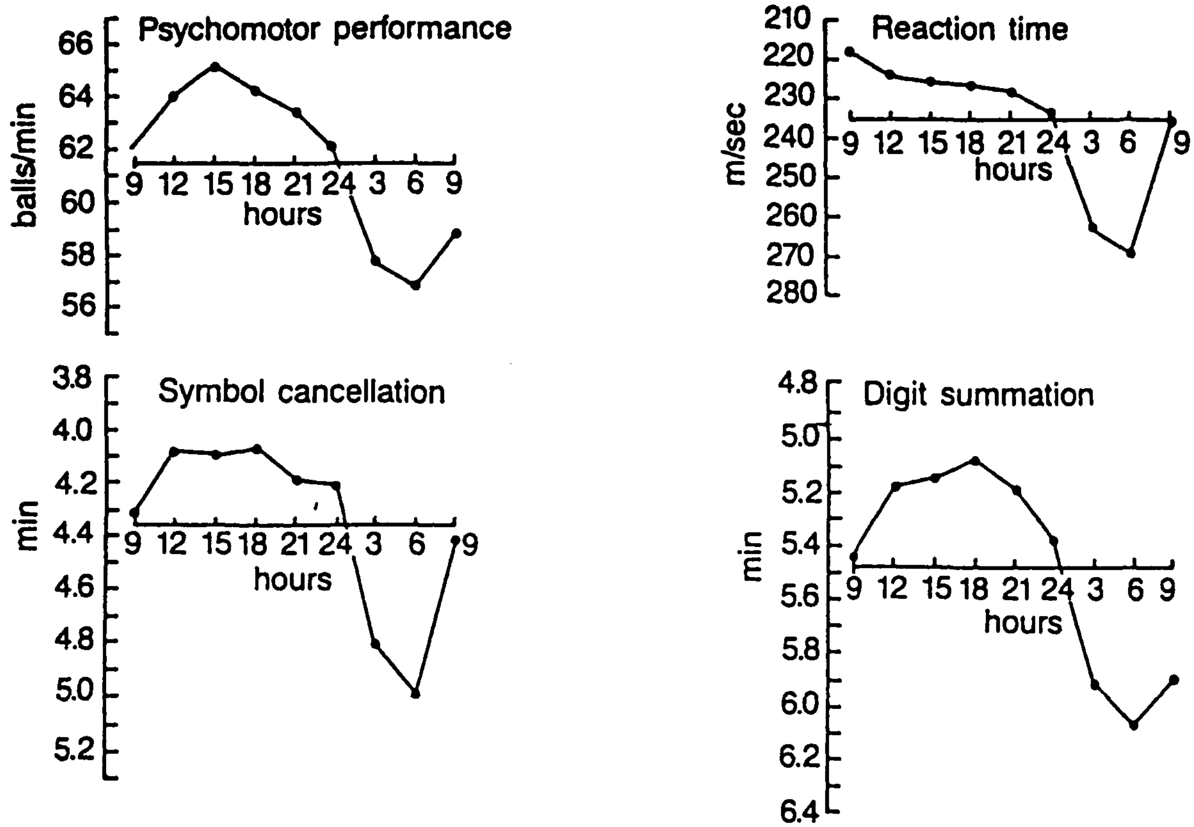


Figure 8.13. Graphs showing how human performance varies during the day with a rhythm corresponding to body temperature. (from Klein et al., 1972)

extra sleep in advance of a period of time when you are going to miss a lot of sleep. It can do a very good job of compensating for the later loss of sleep.

A second finding from the short-haul study was that sleep loss each night is greater on layovers than at home. Generally, the pilots were sleeping less per night on the layovers. The sleep was also more fragmented during the layovers. Graeber also examined the buildup of fatigue across the four days of flying, and found that this buildup (measured by the pilots' subjective rating of how tired they were), was really greatest after the first day of the trip, with a more modest increase in fatigue after the third and fourth days.

Now consider what each day of the trip is like. Some days are very fragmented and consist of three or four different legs on different aircraft -- up to seven or eight takeoffs and landings at different airports. Other days may involve only one flight with a fairly long layover. Thus we can distinguish between busy days and relatively nonbusy days in terms of takeoff and landings. Graeber's third conclusion was that sleep was better following a busy day than following a relatively light day. That is not altogether surprising. The busier the day, the more takeoffs and landings, the more fatigue within a day, and, therefore, the better the sleep will be after that day is over. A fourth conclusion from Graeber's study is that ***down-line changes of schedules are bad for sleep planning.*** If, after the second or third day into the short haul, the pilot was informed of a sudden change in the flight schedule, this change seriously disrupted the pilot's sleep schedules. It was almost as if the crews could preprogram themselves for how much sleep they were going to need each night into the short haul. However, if that schedule was suddenly disrupted by a change, that change disrupted the preprogramming. For pilots who have done operational flying for commercial airlines, most of these conclusions are probably not surprising. The important point is, for the first time, they are firmly documented in an objective study with data.

The second major component of Graeber's work was a study of long-haul flights. These are transoceanic flights that typically involve time-zone changes of six or more hours. To understand the effects of those long-haul flights, we need to consider a little bit more about this natural circadian rhythm. It turns out that the period of the natural rhythm is not exactly 24 hours, but it is actually about 25 hours. Studies of people who have gone into caves where they have no sense of waking in the natural day/night cycle reveal that these subjects tend to adopt a 25-hour schedule rather than a 24-hour schedule. There are interesting reasons why this is the case, but it is very clear that our natural schedules tend to be longer than the daylight forces us into. When left to our own devices during the week, we tend to stay up later and later each night, and we tend to be late stayers more than early risers. What happens, nevertheless, when we go into a long-haul flight is that we have suddenly moved to a situation where the day/night cycle in the environment where we land, is different from the day/night cycle that our brain has adapted to when we took off. This phenomenon is called ***desynchronization.***

Desynchronization is represented by Figure 8.14. The upper graph represents the westbound flight and the lower graph represents the eastbound flight. The dotted line is the natural circadian rhythm that was formed when we left our home base. So it is the same no matter whether we are flying west or east. The solid line for the west- and eastbound flights is the circadian rhythm at the destination. As we fly west, we are flying with the sun, and initially undergo a

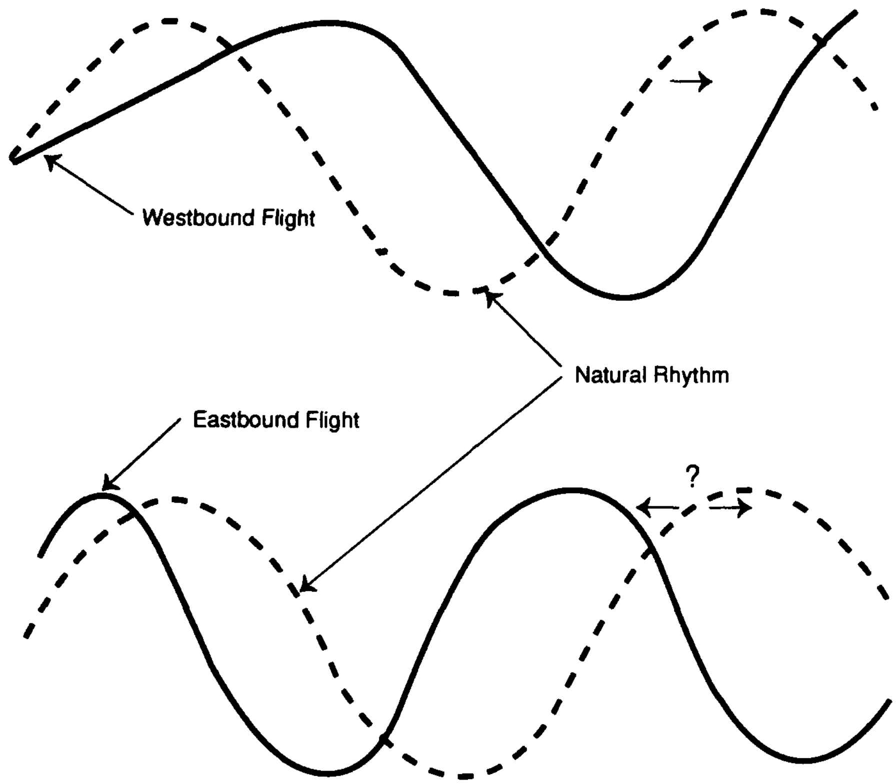


Figure 8.14. Graphs showing desynchronization on east- and westbound flights across time zones. (original figure)

very long day. As we reach the new destination, now we have a day/night cycle, but it is shifted ahead of what our natural cycle is. So when our brain thinks it's night, it is still afternoon. When we are flying east, on the other hand, we have a very fast day initially. When we reach our destination, again there is desynchronization. Now when our brain thinks it is night, it is morning. The data in either case obviously suggest that there is a mismatch between our circadian rhythms and the post flight day/night cycle.

The data also suggests that it is considerably easier to adapt to westbound flights than eastbound flights. When flying west, the natural rhythms have an easier time lengthening themselves to get in synchrony with the local day/night cycle. On the other hand, when flying east, it is as if the rhythms don't know whether to contract and make a very short day, or expand to make a doubly long day. There are, indeed, some reliable data indicating that the eastbound flights, which condense the day, are worse than the westbound flights which stretch the day. These data come, in part, from examining the way in which different characteristics of the physiological systems adapt to the new rhythms. In other words, you have got a natural rhythm which was in existence when

you left, and you acquire a new rhythm which you should take on when you reach your destination. The longer you stay at your destination, the more the old rhythm is going to shift into phase with the new rhythm. We can then plot how rapidly that shift takes place.

Table 8.2 shows the shift rates for different variables after transmeridian flights, either westbound or eastbound.

Table 8.2
Shift Rates after Transmeridian Flights for Some Biological and Performance Functions

	Westbound	Eastbound
Adrenaline	90	60
Noradrenaline	160	120
Psychomotor performance	52	38
Reaction time	150	74
Heart rate	90	60
Body temperature	60	39
17-OHCS	47	32

(from Klein et al., 1972)

The numbers in the table are expressed in terms of the amount of shift in minutes per day, so that a higher number indicates a more rapid shift. What you see is that generally the numbers for the westbound flights are higher than the numbers for the eastbound flights. In fact, sometimes the westbound shifts are as much as two times faster than the eastbound. The table shows the rate of uptake of adrenaline and noradrenaline, psychomotor performance and reaction time, heart rate, body temperature, and a body chemistry measure (17-OHCS). Each of these different rhythms seem to shift at a slightly different rate. Therefore, in transcontinental or transoceanic flight not only is your rhythm out of synchrony with the rhythm of day and night at your new destination, but all of your different rhythms are out of synchrony with each other because of the different shift rates. Thus there is kind of a "double whammy" to readaptation. Different things are lost at different times, and different things are regained at different times.

The last conclusion of the long-haul flights study is that the return to normal is a relatively gradual one that on the average takes about four to five days before the new rhythms regain synchrony with the local environment. This figure is

probably more like five to six days after an eastbound flight, and perhaps three to four days after a westbound flight. Figu_ 8.15 shows some more data

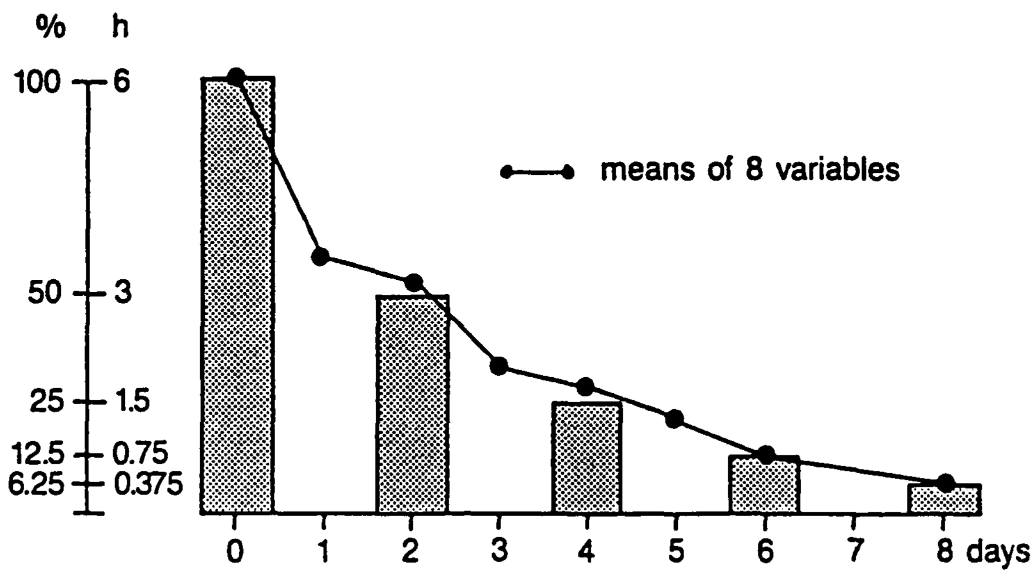


Figure 8.15. **Average resynchronization of variables for eight post-flight days. (from Wegmann et al., 1986)**

representing this shift. It shows how much resynchronization took place for different variables (body temperature, performance, etc.) after the first through the eighth day. Notice that even after eight days, subjects still haven't completely resynchronized with the new rhythms, although most of the resynchronization took place after the second and third day. The bottom line question of course is whether this desynchronization leads to a higher number of pilot-induced accidents or poorer pilot performance. At this point, there isn't a good database to suggest that is the case. In other words, there aren't accidents that have been directly attributed to the resynchronization problem, but there are certainly suggestions that it may have been a contributing cause in some instances.

Recommendations

There are a number of recommendations that have come out of the research on sleep resynchronization, and these are, again, taken from Graeber's work (Graeber, 1988, 1989). His chapter recommends that pilots should sleep when it is most effective and do so within the natural cycle. Where possible, sleep ought

to be scheduled at late night, early morning hours in the phase with the rhythms to which the body is accustomed. Extra sleep, rather than deprivation prior to a short haul, is advised. Following a long transoceanic flight, Graeber argues it is better not to sleep immediately after one's arrival, but simply try to stay awake until the local bedtime, particularly if one is going to be adjusting for some time to new rhythms. During any 24-hour period, sleep is relatively more effective before takeoff than after landing during a layover. So following a landing, sleep during this period is going to be better just prior to the subsequent takeoff. This is consistent with the idea of prophylactic sleep, sleeping in advance of a period where one knows sleep deprivation is likely to occur. Prophylactic sleep is helpful and much more restorative than sleeping just after a period of time without sleep.

A somewhat more controversial issue, but one that is certainly receiving some research interest, concerns ***controlled napping***. How effective is controlled napping in flight, assuming, obviously, that somebody else is awake at the controls. The studies that have been done of napping indicate there are really two sorts of napping. First, there is ***micro-sleep***, where one may doze off for a couple of seconds or a very short period of time. There is very little evidence that micro-sleep, in itself, is effective in restoring sleep loss. Then there is a bonafide nap. There is a minimum amount of time, about 10 minutes, before a nap can be effective in terms of restoring some sort of sleep loss.

Another phenomenon that relates to naps is the concept of ***sleep inertia***. It's something that is intuitively familiar to all of us. Sleep inertia describes the cognitive inertia we experience immediately after waking up. In fact, for 10 minutes or so after one wakes up, there is an inertia that inhibits our ability to respond quickly, think fast, and so forth. This is well-documented in the research of Chuck Czeisler at Harvard, which suggests that any program of controlled napping has got to be one in which the wake-up time is well in advance of the time one may have to carry out some sort of high-level cognitive activity or rapid action. If this is applied to a pilot flying transoceanic, you don't want to wake up just before you start making the important decisions required on the approach, but rather with sufficient time to dissipate that sleep inertia before such decisions are required.

In conclusion, it should be noted that the findings and recommendations reported here result from pooling information from a lot of data sources, many of them not taken from aviation. Furthermore, the causal links between the different forms of sleep disruption and pilot error have not always been conclusively established. Nevertheless, it is prudent to believe that there are some direct implications to aviation performance.

Human Error

Anytime one talks about human error, there is a tendency to do a lot of finger-pointing. Pilot error comes up with a red flag as being a frequent cause of a disaster or accident. Training in engineering psychology, however, leads one to conclude that when errors do occur, they rarely occur as a result of a mistake made exclusively by the pilot. Typically errors are caused by some training-induced, schedule-induced, or design-induced factor that made that error almost an inevitable consequence--something that was bound to happen sooner or later. This is actually a positive philosophy, for it suggests that there are usually steps that can be taken to reduce the likelihood of error.

A number of studies that have looked at pilot errors have tried to categorize the nature of the various errors in terms of where they occurred, how they occurred, and what they were the result of. The approach to pilot error classification that is consistent with the information processing model presented in Chapter 7 is one that identifies four major kinds of errors. In this model of information processing, there are the stages of perception and understanding the situation (situation awareness or diagnosis), formulating some intention for action, (deciding what to do about it and making a choice), and finally executing the action. When taking an action, we often rely upon our memory, both short- and long-term, to help us recall the rules of what it is we are supposed to do. Within this context, two researchers, Norman (1988) of the U.S., and Reason (1990) of the U.K. have come up with similar ways of classifying errors. Classification is important because the different kinds of errors seem to have different remediations, or different fixes. This classification is nicely applied to aviation in Nagel's chapter in Wiener and Nagel's book on *Human Factors in Aviation* (Academic Press, 1988).

Categories of Human Error

In Reason and Norman's Classification scheme, there are, first of all, what are called mistakes, a misunderstanding of the situation. ***Knowledge-based mistakes*** occur when you don't have the knowledge to understand what is going on. ***Rule-based mistakes*** occur when you select the wrong rule to make a decision. Forgetting is another type of error. You forget what is going on, what mode you are in, and you make a mistake. You have ***lapses***, where you simply forget what you are doing and therefore do the wrong thing. Finally, you have errors of the execution of action, which we call ***slips***. A slip occurs when you know what to do, but you slip and do the wrong thing. You hit the wrong button on the control display unit, for example.

We can represent these different types of errors in terms of different characteristics of a pilot's behavior. A knowledge-based mistake might be a misdiagnosis when the pilot doesn't understand what is wrong with an engine. A rule-based mistake would characterize the situation when the pilot knows what is wrong, but chooses the wrong action. The pilot realizes that an engine is malfunctioning, but intentionally reduces power to the engine, rather than shutting it down completely. With a slip, the pilot intends to perform the correct action, but simply executes it incorrectly. For example, the right engine is known to be failing, and the pilot intends to shut it off but shuts off the left one instead.

Considering these error types in more detail, knowledge-based mistakes typically result from inadequate knowledge, usually a consequence of insufficient training or the inadequate or confusing display of information. A good example of a knowledge-based mistake would be misinterpreting flight path information and ground-based features, and landing at the wrong airport. Somehow your knowledge and interpretation of the available information is simply wrong, and you have made a mistake about where you are. Knowledge-based mistakes often occur when attention is directly focused on the task in which the error is made. The pilot who lands at the wrong airport typically doesn't do so because of failure to pay attention to where he or she was going. In fact, the pilot is usually paying fairly careful attention to the aircraft's course at the time, but is simply confused. Knowledge-based mistakes often occur at times of very high working memory load. The operator is usually in a state of uncertainty and hesitancy. Finally, the detection of knowledge-based mistakes is often very slow. As a consequence, you often don't realize the mistake was made until it is too late. These are often characteristics of diagnosing system failures. The pilot is focusing a lot of attention on the demanding diagnostic task. Some human error analyses have been carried out in the domain of nuclear process control. One study looked at 80 process control errors committed in actual plant operations and found that out of the 80 errors, half of them were knowledge-based mistakes. The operators were never aware that they made any of the mistakes. They always thought they made the right decision until the consequences were felt later on. The main remediations for knowledge-based mistakes are (1) training, thereby giving people better knowledge, and (2) displays that provide operators with better, more integrated information.

Rule-based mistakes also result from inadequate knowledge. The diagnosis is correct, one may know the correct status of the world, but one's decision of what to do about it is wrong. It is as if the pilot has a rule of thumb of what to do in case of failure X. Failure X is correctly diagnosed, but the rule is wrong, and therefore the wrong corrective action is carried out. Rule-based mistakes occur when attention is highly focused on the task. Once you diagnose

the situation, you act with a high degree of certainty, even though you are acting incorrectly. As Reason says, one's actions are "strong but wrong." Training is one antidote. Automation assistance is also a possible aid in lessening the likelihood of rule-based errors. It can provide some guidance, given a certain kind of diagnostic condition, of what the appropriate rule to be followed should be.

Two kinds of memory errors have been referred to. One of these, more common in computerized systems, is called a ***mode error***. A mode error occurs when the operator forgets the currently active mode of operation. The simplest example is the typewriter or computer keyboard. Suppose you are typing along and you press the CAPS LOCK key, that makes everything you type in capital letters. Then you forget what mode you're in, and start typing digits. On the conventional typewriter keyboard instead you will get: $&&@#(&! This is a mode error. Mode errors are likely to occur in any multimodal system in which the same response can generate various results, depending on the mode setting. Mode errors are not likely to occur if the operator is new at the system, and is concentrating very intensely on remembering what mode the system is in. The more familiar you get with the system, the more you stop paying attention to what mode you are in, and the more likely you are to make a mode error.

As we deal with automation devices that are increasingly based upon different modes of operations, like multimode autopilots, mode errors are likely to occur with increasing frequency. The remediation for mode errors is to provide very strong reminders of what mode of operations one is operating in. Consider, for example, multiple modes of autopilot control where the level of guidance is controlled by a wings leveler or heading control. There should be something highly visible and continually available to remind the pilot what mode the system is operating in. Another remediation for some mode errors in computer operations is simply to use dedicated keys or one-to-one mappings between key and function. This means you press a key and it always does only one thing. This feature avoids a design where a given key can activate very different functions depending on the mode setting of some other key. However, it is often a more economical design to have multimode keys rather than one-to-one mapping as far as space is concerned.

A second form of memory errors are the occurrence of ***lapses***. Lapses result whenever a procedure is forgotten. One simply forgets to do something in a series of steps. Lapses often occur when a long series of actions are required to reach the goal. This is obviously the case in many checklist operations, like pre-takeoff, pre-landing, etc. Lapses are more likely to occur when a procedure sequence is interrupted, then later resumed. Perhaps in following a set sequence

of A, B, C, there is some interruption; later on the operator jumps back in and forgets that step D was not yet performed and goes right on to E, F, and G. The National Transportation Safety Board (NTSB) report of the Northwest Airlines crash in Detroit, in which the flaps were not deployed, indicates that a lapse was a very likely cause. The pilot was going through the taxi checklist on the runway. Then there was a series of disruptions by air traffic control requesting a change in the runway. Investigators inferred that somehow that checklist was resumed, but that one critical step of deploying the flaps had been left out. Other contributing causes to the disaster were, of course, also identified. There were a number of fail-safe operations that did not work and thereby allowed the error to occur. Many of these fail-safes were also related to automation, but it is clear that the checklist procedure contributed a major potential source of error. A remediation of this kind of situation would be a checklist design which avoided forcing pilots to go through multistep sequences that do not have a clear ***prompt*** that guides them through the checklist, saying "do this, do this, do that, do the other, check this, check that." Even with such external prompts, there is no guarantee that the steps will all be done, but it certainly is an important safeguard. Degani & Wiener (1990) have written a nice summary of the human factors of pilot checklists.

A ***slip*** is an error which occurs when you have diagnosed a situation correctly, you have formulated the correct intention, your rules of what to do are correct, but somehow there is an incorrect execution. The error category of slips sometimes includes mode errors and lapses. You either left out a step or did an extra step. One example of a slip is hitting the wrong key on a keyboard. Another example is grabbing the orange juice instead of the syrup, and pouring it on your pancakes. Certainly in aviation, there are lots of situations where the wrong control has been activated. The pilot may activate the flaps rather than the landing gear, when the pilot surely knows the landing gear and not the flaps is what should be activated. What are the conditions that cause slips in the first place? There are really three triggering conditions. First of all, a slight deviation from the most expected or frequent behavior sequence is intended. There is a familiar pattern of activity you carry out most of the time, and the needed pattern is similar, but slightly different. Second, the conditions or location and the feel of the intended action is similar to the conditions of the less frequent action. So most of the time you are doing A, B, and C. Under these circumstances, you plan to do A, B, and C', which is slightly different than C. It may be a slightly different control, a control located close by, but in a slightly different location to the normal control C, or a control pulled upward (C'), instead of downward (C). A third triggering condition for slips is that performance in carrying out the sequence of actions is fairly automated, so attention is usually directed elsewhere.

A general characteristic of slips is that they are "strong but wrong." An operator commits to the action, and usually does it with the same degree of certainty as the correct action. Fortunately, we are usually fairly good at detecting our own slips just as they are made. As we type or enter data into a CDU, it is very obvious when we make a slip, as if the finger knows before the brain knows that it has gone to the wrong place or setting. With a particular switch in an aircraft, you may know immediately that you made the wrong choice.

The fact that we are good at catching ourselves making slips has some important implications for how we remediate them. Remediation of slips is a major issue in system design. Since slips usually occur when attention is directed elsewhere, that means slips usually occur on sequences of behavior that are fairly well learned for operators that are highly trained. So remediation is really not so much in training as it is in system design--remediation includes such things as avoiding the design of similar controls with similar physical actions which must be used in similar conditions. Good design avoids circumstances where you have two similar switches that are flipped in similar conditions but for different purposes. Always try to adhere to SR compatibility. One of the major culprits causing slips is the incompatible response mapping, discussed in Chapter 7. Here without paying attention, the pilot may have a tendency to move something in the wrong direction because the right direction was an incompatible response.

Error Remediation and Safeguards

In this section we review and present a series of recommendations that psychologists have proposed to remediate human error -- eliminate it, or reduce the likelihood of its unpleasant consequences. First, there is the issue of allowing for ***reversibility*** of actions. Such an allowance creates what we call a ***forgiving system***. As we noted, operators are usually pretty good at monitoring their own performance and detecting their own errors if there are slips. Once you've made an error, it is nice to have a chance to correct it. Some systems have an "error capture" mechanism, which captures and delays the response a little bit before its consequences can effect the system. That's not always a feasible design option, but there are situations in which it can be made feasible. There are computer systems that, whenever you press a button that involves deleting a major file, will come back with a message that says, "Are you sure you want to delete this?" That is like capturing your behavior before it gets passed on to the system. Slips often involve throwing things away. Don Norman, the author of *The Psychology of Everyday Things*, stores all of the trash baskets in his office for 24 hours in a separate room before they are emptied. If someone in the office realizes the next day that they inadvertently threw out something important, they can go into the room and pull out the information.

This is a forgiving system. On the other hand, if, on an airplane, you slip some paperwork into the seatback pocket, then forget it when you exit from the plane, your chances of getting it back are slim. As soon as the plane is empty, the maintenance crew will have almost immediately cleaned out the seatbacks and destroyed it. That is not a forgiving system that acknowledges the fact that people do have lapses of this sort.

The idea of reversible actions, or forgiving systems, where a slip can be reversed and undone before it is passed on to the system has led to a philosophy of human error that is somewhat of a marked departure from an earlier philosophy. That earlier philosophy was that human errors are bad, and whenever they occur, we ought to try to remediate them. Therefore, we ought to try to redesign the system to make sure an error doesn't occur in the first place. This philosophy has led to two approaches. One is called "bandaids." In the bandaid approach, the system gets more and more complex, because every human error is a cause for another design feature (i.e. a bandaid) that tries to eliminate the human error. This correction, by making the system more complex, very often creates conditions conducive for another error (mistakes become more likely with more complex systems) and doesn't acknowledge the fact that errors are probably always going to happen to some extent in any case; any fix for one sort of error may be likely to produce another error. The second approach characterizing the old philosophy that all human errors are bad is one which pushes automation as an ideal because of the belief that a computer can perform better than a human if there is a mistake. The problem with automation is that the designer is usually transferring the responsibility for human error to someone else. For example, this responsibility may be transferred from the pilot to the computer programmer who is just as likely to make the errors as the pilot.

In contrast to the earlier philosophy, the proponents of forgiving systems make two assertions about errors. They say that an error is, first of all, unpredictable and inevitable. No matter how we design the system, and patch it with bandaids, errors are always going to occur to some extent. Furthermore, they say that error is sometimes a necessary consequence of the fact that the human is a flexible performer. It is that very flexibility that makes us want to keep humans involved in the first place. Pilots have flexible problem-solving skills, and that's good. There is an inevitable cost to that flexibility, and that sometimes is going to lead to the wrong action in inappropriate circumstances, but we still want to maintain that flexibility because of its positive qualities. We have to accept the consequences, which are the occasional errors; therefore, our philosophy of redesigning the system should be one that says errors are going to occur but let's design the system in a way in which they can be tolerated. This is the philosophy for ***error tolerant systems***.

In this vein, Earl Weiner has discussed the concept of the ***electronic cocoon.*** The idea here is that a pilot ought to be free to make a lot of different responses, some of which may be incorrect. The appropriate role of automation would be to simply monitor the performance envelope of the aircraft, and only intervene if the errors are serious enough to bring about a serious consequence. The idea is to have some master computer monitoring the pilot, but allow the pilot a lot of opportunities to make errors and to correct them before things get bad. Bill Rouse and his associates have done a lot of work on this concept for the Air Force, as part of the Pilot's Associate program, designing electronic copilots that can monitor the pilot's performance and act as a cooperative crew member. Their concept is that of an intelligent system which can monitor human performance and infer the intentions of the human control actions. You have a pilot interacting with a task under intelligent monitoring. The pilot's behavior is providing information to the monitor. The monitor, in turn, can take a series of actions in the face of the pilot's behavior, if the monitor detects that the pilot might be making mistakes. Rather than just simply taking over for the pilot, Rouse and his colleagues suggest that this intelligent monitor might go through a hierarchy of guidance. At the very first level, if the intelligent monitor infers that the pilot is doing something that is amiss, it might do nothing more than increase vigilance. If there is continued evidence that the pilot's behavior is inappropriate, the intelligent monitoring system might say some things to the pilot, like "*Are you sure you want to do this? Are you watching your airspeed?*" If the error worsens, the monitoring system might prompt the operator with some advice like lowering or increasing airspeed, etc. Only under the most serious error circumstances will the intelligent monitor assume command automatically and correct the error.

Error in a Systems Context

In conclusion, it is important to consider the concept of human error in a much larger domain of overall system integration. Jim Reason has done so by introducing the concept of error as a "resident pathogen." Reason speaks of a "latent error" or resident pathogen as a virus that sits in the system not causing any particular abnormality, but waiting for some conditions to trigger it. Reason examined a lot of different case studies of major disasters such as Chernobyl, Three Mile Island, the Bhopal incident at the Union Carbide plant in India, and the sinking of the ferry boat "Herald of Free Enterprise." This was the ferry boat that sank crossing the English Channel after the captain left the loading door open in heavy seas. The boat filled up with water, sank, and scores of lives were lost. All of these were disastrous events that were directly attributable to operator error at some final point in the chain of events. However, Reason concludes that, in fact, the operating conditions in these complex systems were conditions that were poorly designed with a potential error lurking there

somewhere (like the pathogen). All that was needed was for one operator to "trigger" the system, and make these inevitable errors occur. Furthermore, he argues that there are a large number of potential causes of these catastrophes within complex systems. Rather than pointing a finger of blame at a particular operator who commits the final triggering error, Reason argues that the real remediation should be accomplished by considering a number of ***mediating factors*** that made the disaster a nearly inevitable consequence of a triggering human error.

One of these factors is the collection of hardware defects related to poor human factors concerns of design, construction, and location. System goals that are incompatible with safety also contribute to errors. Very often in industry, system goals are designed towards production rather than safety. These two goals are not always totally compatible. Poor operating conditions have a tremendous impact on the extent to which the goals are or are not compatible with safety. Inadequate training is another factor. Just checking off a box and saying somebody has been through the simulator is inadequate. Poor maintenance procedures is an additional factor that creates conditions for error. The Three Mile Island disaster was a case where maintenance procedures were sloppily carried out, and it wasn't clear to the control room personnel on duty what systems were and were not in operational status. Finally, management attitudes (or lack of guidance) can lead to violations by operators that will help propagate unsafe acts. The operators at Chernobyl provided a nice example of where the people at the plant simply do things that they knew weren't supposed to do, because the guidelines had said it was all right to do so. We are all making violations every time we surpass the speed limit. We know we are going over the speed limit by a few MPH, because we don't have incentive not to do so.

Reason's final point is that sometimes even though a system is very well designed from a human factors point of view, following the sort of prescriptions we have discussed here, there will still be human errors because of the failures at all of these other levels. This is a systemwide approach to human error.

Chapter 9

Cockpit Automation

by Richard F. Gabriel, McDonnell-Douglas, retired

Introduction

The Federal Aviation Administration (FAA) has a direct and pervasive influence on aircraft design through its certification process, and on operations through its design and operation of the Air Traffic Control System (ATC). In spite of the FAA's broad regulatory administrative role, it is difficult for rules and regulations to keep pace with rapid technological advances in aircraft design and operation. It is therefore important that FAA personnel have an understanding of the impact that advanced technology (automation) may have on those who operate these systems, so that the benefits of automation can be realized without unacceptable side effects.

In recent years, increasing levels of automation have shaped and changed the aviation industry. These effects include:

- Economic impacts - growth in passenger demand, increase in fuel prices and other operating costs, increased competition among airlines;
- Changes in airspace and airport configuration - capacity limitations, hub-and-spoke concepts, air traffic control requirements;
- Effects on equipment - increased equipment reliability, increase in aircraft longevity, aircraft design and performance improvements, increased automation of flight decks;
- Effects on operators - reductions in crew size, reduced emphasis by airlines on training, changes in crew qualifications and availability.

This review will consider the human factors issues of automation from the operator's standpoint. Although the discussion is relevant to ATC as well as flight crews, emphasis will be on cockpit applications.

Definition

Automation has been defined as the incorporation or use of a system in which many or all of the processes...are automatically performed by self operating machinery [and] electronic devices. (Webster's New World Dictionary, 1970). Figure 9.1 depicts the progress of automation in aircraft and indicates automation has been increasing since the origin of heavier-than-air flight. Automation is not an all-or-nothing proposition. Sheridan (1980) has identified ten levels of automation, from totally manual (100 percent human controlled), to systems in which a computer makes and implements a decision if it feels it should and the human may not even be informed (100 percent computer controlled). Current systems generally fall between these extremes, but the trend is to reduce the role of the human and move away from human control even in decisionmaking. Self-correcting systems are becoming commonplace in newer aircraft. Table 9.1 presents Sheridan's levels of automation.

Summary of Aviation Automation Concerns

Some human factors specialists have expressed concern about designers' overreliance on automation to perform flight functions. Recent developments--particularly the availability of small, powerful digital computers--have led to systems designs that not only control the aircraft for much of its flight, but may even replace crew decision functions in the hope of reducing human error. An

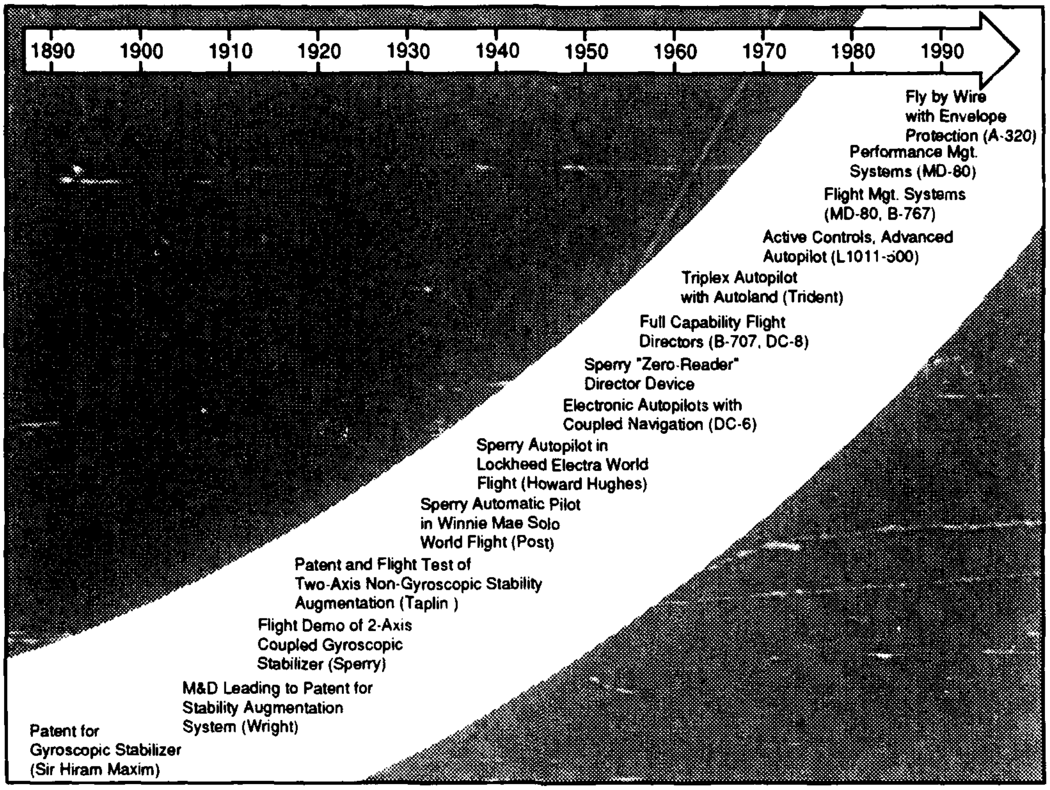


Figure 9.1. A Timeline of the Development of Aircraft Automation.

example is envelope protection, in which certain maneuvers such as a stall cannot be induced by the crew either intentionally or unintentionally.

Another concern is the possibility that even redundant systems may fail. In these situations, flight crews may experience difficulty in diagnosing problems and performing corrective actions if they have been lulled into overconfidence by highly automated flight systems, or have lost the fine edge of their skills as a result of disuse.

Some of these automation concerns are illustrated by the following scenario:

> A pilot of average skill is captain of an advanced, highly automated aircraft incorporating features such as relaxed static stability, full-time augmentation, a sophisticated flight guidance and control system, and "envelope protection" with most failures detected and corrected. The captain has flown in this type of aircraft for some years and has recently upgraded to his position. The crew flies in the automatic modes most of

> the time. They are making an automated approach and landing, when, at the middle marker, a major electrical failure causes the aircraft to revert back to its basic characteristics. The crew has to take over control of the aircraft, make the correct decisions, and take appropriate actions. Additional factors may complicate their decisionmaking: night, bad weather, the start of a bid cycle, fatigue and other plausible and realistic influences.

The ultimate question for designers, manufacturers, operators, and certifiers is whether safety will be enhanced by incorporating a specific automated capability. The answer lies in the crew's ability to interact with the automated system effectively and to take over in the event of a failure or a situation not foreseen by the designers.

Table 9.1
The Spectrum of Automation in Decision Making (Sheridan, 1980)

100% HUMAN CONTROL	1.	Human considers alternatives, makes and implements the decision.
	2.	Computer offers a set of alternatives which human may ignore in making decision.
	3.	Computer offers a restricted set of alternatives, and human decides to implement.
	4.	Computer offers a restricted set of alternatives and suggests one, but human still makes and implements the decision.
	5.	Computer offers a restricted set of alternatives and suggests one, which it will implement if human approves.
	6.	Computer makes decision, but gives human option to veto before implementation.
	7.	Computer makes and implements decision, but must inform human after the fact.
	8.	Computer makes and implements decision, and informs human only if asked to.
	9.	Computer makes and implements decision, and informs human only if it feels this is warranted.
100% COMPUTER CONTROL	10.	Computer makes and implements decision if it feels it should, and informs human only if it feels this is warranted.

Experience with Automation in Nonaviation Systems

Experience gained in the design and operation of automated systems in non-aviation environments is often relevant for aircraft systems. Process plants such as power plants, oil refineries, factories, and offices have adopted various levels of automation.

Nuclear Power Studies

Designers of nuclear power plants have incorporated many automated features in their control systems to avoid catastrophic human error. This is because they fear that the human operator may not be able to respond to system emergencies that occur with manual systems. They believe that automation, because of the complexity of nuclear power plant design and processes, can solve this problem.

Yet, it has been found that automated safety systems aren't necessarily the answer. An evaluation of 30,000 nuclear plant incidents revealed that 50 percent occurred through unique combinations of machine and human error (Woods, 1987).

The Three Mile Island accident is a case in point. The initial blame for this incident was assigned to humans. Investigators found, however, that the design of the human interface was greatly deficient. Designers had not considered the human functions systematically. They had paid little attention to display/control design or work station layout. The control room was filled with banks of almost identical controls and displays that made it difficult to identify the appropriate information source or required response to system problems. For some functions, the operator could not see the display and the corresponding controls simultaneously. To compound these problems, training of control room operators had been inadequate.

After the Three Mile Island accident, the response of managers and designers was to further divorce the human operator from system control through even more automation. Extensive programs for redesigning the displays and controls were initiated. One involved changing the warning system from a tile (e.g., legend light) system to a computer-based system. The purpose was to automate the alarm system and reduce display clutter. The result was disappointing. The computer-based system wasn't programmed to anticipate all the possible combinations of events that could occur; the operators lost the ability to integrate display information by recognizing patterns of lights and thus gain insight into the fundamental problem.

Additional multimillion dollar studies were initiated in several countries to develop a computer-based fault diagnosis system. The goal was to reduce the operator's role in fault diagnosis. It was found that fault diagnosis could not be totally automated. The computer solved the easy problems, but the tough ones were left for the operator. The operator tended to be overloaded with data and also tended to be deskilled; that is, he or she lost the value of practicing on the easy problems. Ultimately, the effort to automate fault diagnosis was abandoned.

Office Automation

Research in office automation has shown that no system, even a very simple one, is ever completely defined by designers. One reason is that the system is not always used for the purpose initially intended. A screwdriver offers a simple example. It was designed to drive or loosen screws. But it is also used to open lids of cans, scrape surfaces, clean fingernails, and even as a weapon. Similarly, a wire coat hanger may be used to help open a locked car.

The same variability in application is found with automated systems. Inventory systems may be used differently as business grows, shrinks, and/or conditions change. Accounting systems may have to be altered as tax laws change. Even office electronic mail systems may be used variably as security needs or capacity requirements change (Card, 1987).

According to articles in the public press, many of the increases in productivity anticipated from office automation have not been realized. Moreover, the costs in personal satisfaction and well-being have been high. Worker motivation has suffered as jobs have been changed and depersonalized.

Table 9.2 offers some conclusions various authorities have reached after studying automation in arenas other than aviation.

Table 9.2
Conclusions Based on Research in Nonaviation Automation

- Humans tend to be less catastrophically affected than computers when subjected to severe overload (Sinaiko, 1972).
- Human performance is degraded when automated systems perform very well (Rouse, 1977).
- In situations that require strict vigilance, information sampling and transfer is done better by humans than by automated systems (Crossman, Cooke, and Beishon, 1974).

Table 9.2 (Cont'd)
Conclusions Based on Research in Nonaviation Automation

- About 20% of human input errors go undetected (DoD, 1985).
- Automation can lead to sloppiness (Card, 1987).
- The nuclear power industry's evaluation of 30,000 incidents revealed that about one-half occurred through unique combinations of machine and human errors. Trying to change the automated system sometimes created new difficulties because the interaction between humans and machines was changed in unforeseen ways (Woods, 1987).
- Automated systems usually solve simple problems but fall down in more complex cases (Roth, Bennett, and Woods, 1987).
- We need to complement the design for prevention of trouble with the design for management of trouble (Roth, Bennett, and Woods, 1987).
- Computer systems should be designed as a tool, not as a replacement for the human (Roth, Bennett, and Woods, 1987).

Experience with Automation in Aviation

The effects of automation on human performance are difficult to assess. Many of these, especially boredom and loss of skill, occur fully only after extended periods of time. To quantify the effects of automation on human performance would require time-consuming longitudinal studies under controlled conditions. There are sources of data, however, such as opinion surveys and accident-incident data that can help provide insight. Some of these data will be summarized and discussed in this section.

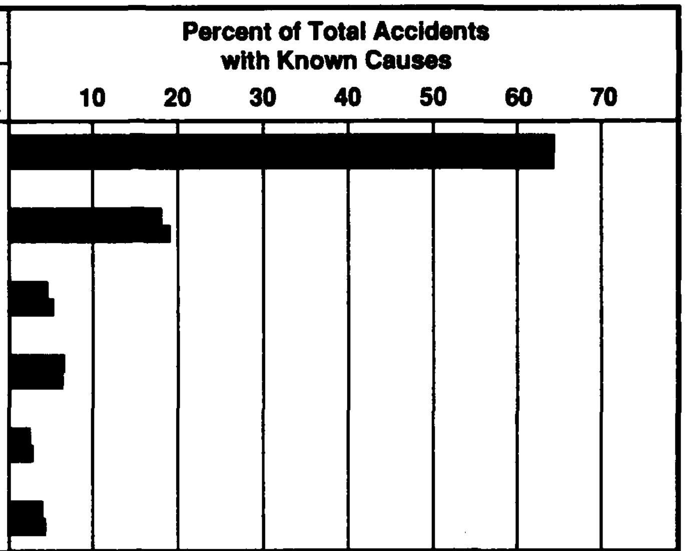
Accident Data

Errors on the part of the flight crew have historically been cited as a primary cause in most accidents. Figure 9.2 presents data tabulated by Boeing Commercial Aircraft Company and cited by Nagel (Nagel, 1989). It shows that flight crews have been identified as a primary cause for accidents about 65 percent of the time. The next largest primary cause--airframe, power plant, or aircraft system failure--accounts for less than 20 percent of accidents.

As shown in Figure 9.2, the flight crew has remained a primary cause of accidents at about the same frequency over the years since 1957. The reason for the overall improvement in system safety is probably not a result of any single factor. Reliability of equipment, better knowledge of weather, and almost universal availability of instrument landing systems have undoubtedly contributed. The largest gain in safety of air travel was made during the 1977 - 1981 period. (This was before the introduction of third generation jets that dramatically increased automation in the cockpit.) The following period

Worldwide Commercial Jet Fleet

Primary Factor	Number of Accidents: Total	Number of Accidents: Last 10 yrs.
Flight Crew	441	150
Airplane**	119	42
Airport/ATC	33	11
Weather	32	12
Maintenance	19	6
Misc. (Other)	30	10
Total with Known Causes	674	231
Unknown or Awaiting Reports	73	31
Total	747	262



*Excludes
- Sabotage
- Military Actions
- Turbulence Injury
- Evacuation Injury

**Includes
- Airframe
- Aircraft Systems
- Powerpoint

1959 - 1986

Last 10 Years (1977 - 1986)

Source: Statistical Summary - Boeing from Nagel, 1987

Figure 9.2. Boeing statistical summary of primary cause factors for accidents. (Nagel, 1987).

(during which the MD-80, 757, and 767 were introduced) suggests a slight reduction in safety, but this change may not be statistically significant. Even though flight crew error rate as a cause of accidents has remained constant, flight crew performance probably has improved through better training (use of simulators, for example), better human factors engineering, and other performance enhancements.

The trend in commercial aviation has been toward dramatic improvements in safety. Table 9.3 shows accident trends in terms of the probability that an individual will be killed due to an accident on any nonstop flight in the United States in 5-year increments since 1957, the date jet service was initiated. The data indicate that a traveller is approximately 10 times safer now than in the 1950s.

Table 9.3
Probability of an Individual Being Killed on a Non-Stop U.S. Domestic Trunkline Flight

PERIOD	RISK LEVEL
1957 - 61	1 in 1.0 million
1962 - 66	1 in 1.1 million
1967 - 71	1 in 2.1 million
1972 - 76	1 in 2.6 million
1977 - 81	1 in 11.0 million
1982 - 86	1 in 10.2 million

Incident Data

An incident has been described as an accident that didn't happen--an event that could have resulted in an accident but did not because the crew recovered (avoidance maneuver) or other factors intervened. Since incidents occur more frequently than accidents, they provide sufficient data to identify trends that may allow detection of unsafe conditions and allow corrective measures to be initiated before accidents occur.

The Aviation Safety Reporting System (ASRS) was established by NASA to provide an incident database. The ASRS database includes data from all segments of aviation, including commercial aviation, general aviation, and air traffic control. It is interesting that ASRS incident data presented in Figure 9.3 mirror almost exactly the proportion of human error depicted in Figure 9.2.

A NASA study on classification and reduction of pilot error used the ASRS database to identify problems associated with Control-Display Units (CDU) in cockpits (Rogers, Locan, and Boley, 1989). The CDU is a common feature of automated systems ar' 's a common source of crew errors. It allows the operator to program . .d observe the state of automated equipment. In modern cockpits, it generally consists of a cathode ray tube (CRT) and a related keyboard.

Of the approximately 29,000 reports in the ASRS database at the time of the NASA study, 309 involved CDUs. Table 9.4 provides some specific problems found with CDUs. This analysis of CDUs shows that both human and machine error occurred, with human error predominant. Clearly, humans make errors even in automated systems.

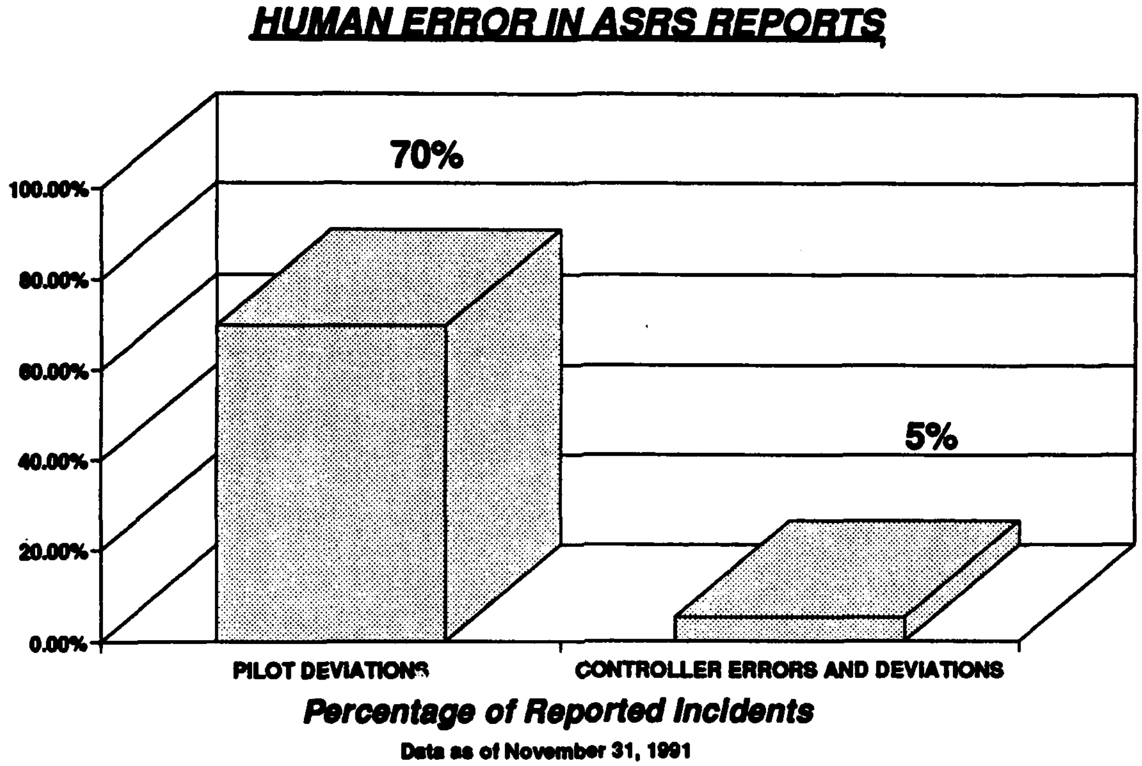


Figure 9.3. Human Error in ASRS Reports.

Table 9.4
ASRS Flight Management System (FMS)/Control Display Unit (CDU) Analysis (Rogers, Locan, & Boley, 1988)

-44 Altitude Deviations	-43 Lateral Deviations
• 19 distracted due to CDU	• 14 incorrect routes programmed
• 8 VNAV disconnect unnoticed	• 10 FMC nav errors
• 7 insufficient time to program	• 8 distracted due to CDU
• 6 crossing waypoints confused	• 7 while in holding
• 4 descent restrictions entered incorrectly	• 3 insufficient time to program
	• 1 runway change

One potential weakness of the ASRS is that reports are voluntary. Not everyone experiencing an unsafe condition reports it. Aircraft equipment malfunctions

probably occur more frequently than are reported to the ASRS, particularly when they do not lead to incidents or near accidents.

However, the FAA requires significant equipment problems to be reported as Service Difficulty Reports (SDRs). An analysis of SDRs for DC-9/MD-80 aircraft during one time period is provided in Table 9.5. This table reveals that of the 445 events included, 201 required crew intervention. Of these, 160 required an unscheduled landing or aborted takeoff. Only four SDRs involved cockpit crew error. In focusing on accidents, it is easy to forget just how significant the crew's role is in averting accidents caused by equipment malfunction.

Data from the Douglas Aircraft Accident/Incident Database supports this conclusion. Table 9.5 shows that of the 736 reports in the McDonnell-Douglas database, 65 percent are related to equipment malfunction. Only 12 percent are related to crew error.

Pilot Opinion

The opinion of the flight crews operating the aircraft provides an important source of information on cockpit design. Although crew opinion is subject to many sources of bias and is not by itself adequate for design decisions, it is a rich source for hypotheses about design advantages, disadvantages, and areas needing intensive study.

Table 9.5
Analysis of DC-9/MD-80 Service Difficulty Reports

TOTAL EVENTS FOR TIME FRAME		445
NUMBERS OF CREW INTERVENTIONS		201
ABORT TAKEOFF	29	
UNSCHEDULED LANDING	131	
EMERGENCY DESCENT	11	
FUEL DUMPING	3	
DEACTIVATE SYSTEM	13	
ENGINE SHUTDOWN	12	
OTHER	29	
AUTOMATIC SYSTEM FAILURES		45
OTHER EQUIPMENT FAILURES		230
CREW ERROR		12*

*Eight of these were related to flight attendants (galley problems, etc.)

NASA performed several field studies of crew acceptance after the introduction of the MD-80 and B-757/67 aircraft. Data sources included direct observation of crew performance on the flight deck during normal revenue service, interviews, and questionnaires. The results reported were generally as follows (Curry, 1985; Wiener, 1985):

- Crews liked automated aircraft.
- There was a slight trend toward reduced workload.
- Late changes by ATC created problems in reprogramming.
- There was a slight trend toward fewer errors with automation.
- As crew experience increased, there was a tendency to turn off the automation (Flight Management Systems) during busy times. Several factors were cited to explain this: mismatch of automated system capability with ATC instructions; slow response of autopilots; problems in crew interfaces; and training inadequacies.

This brief review suggests that many of the same difficulties encountered in non-aviation environments are experienced in automated cockpits.

Reasons Cited for Automating Systems

Operators and manufacturers want to maximize the return on their investment. The decision to invest the huge sums required to develop and certify new systems is not undertaken lightly. For operators and manufacturers to seriously consider automating systems, there must be strong potential for a dividend. For the manufacturer the dividend is increased sales and safety. Table 9.6 lists benefits commonly cited to justify automation.

Most of these reasons emphasize the need for increased efficiency in operation and use of airspace and airports, and improved operations in varying environments. To meet these requirements, designers seek ways of providing lower fuel, maintenance, and crew costs while improving efficiency through higher reliability, greater payloads, and more precise flight path control. Trends in aircraft flight deck design indicate that airline and manufacturer decisionmakers believe that automation will help meet their goals.

Table 9.6
Reasons Cited for Automating Systems

- Enhanced safety
- Reduced human error
- Improved human performance
- Reduced crew workload and fatigue
- Reduced crew training requirements
- Reduced crew size
- Improved efficiency
- Reduced costs
- Increased precision, accuracy, stability
- Performance of functions beyond human capability
- Increased operational capability
- Reduced approach noise
- Reduced weight
- Increased capacity
- Improved passenger comfort and ride quality
- Reduction of boring, tedious, and/or unpleasant tasks
- Improved reliability and schedule performance
- Improved management control
- Improved speed and quality of learning
- Competitive posture
- Reduced task difficulty, more convenience and ease of use

Some Automation Concerns

Designing a new aircraft system as complex and sophisticated as a modern airliner is a formidable challenge, particularly with typical time and budget constraints. Meeting this challenge requires a design team to focus intensely on their objective.

Historically, designers have emphasized hardware development. They have relied on the crews to adapt to their flight deck designs rather than designing cockpits to accommodate the performance characteristics of the crews. Allocating budgets for human factor considerations has been a hard sell for many reasons: lack of understanding, uncertain or unspecified payoffs, undefined criteria, threats to established budgets and schedules, lack of a recognized and/or easily accessed human factors database, and mistrust of human factors practitioners.

It is perhaps ironic that automation--intended to reduce the reliance on humans--may require greater attention devoted to human factors. A British author who has worked extensively in studying automation in process plants

and offices has identified a number of ironies associated with automation (Bainbridge, 1987). Table 9.7 lists some of the author's observations.

Table 9.7
Ironies of Automation (Bainbridge, 1987)

- By taking away the easy parts of the task, automation can make the operator's task more difficult.
- The classic aim of automation is to replace human manual control, planning and problem solving by automated devices. But even highly automated systems need humans for supervision, adjustment, maintenance, expansion, improvement, etc.
- The more advanced a system is, the more crucial may be the contribution of the human operator.
- Designers may view the human operator as unreliable and inefficient, to be eliminated if possible. There are two (2) ironies in this: design error can be a major source of operating problems; and designers seeking to eliminate the human operator still leave him/her to do the tasks which the designers can't automate.
- Efficient retrieval of knowledge from long-term memory depends on frequency of use. (Consider any course which you passed and haven't thought about since.) Knowledge about how to cope with abnormal conditions develops only through use and feedback. Yet the operator is expected to cope with such situations when the reliability of the automated system is the justification for acquisition.
- Current automated systems work because they are being monitored and aided by formerly manual workers. Later generations of operators may not have the requisite skill and knowledge to make the automated system work.
- A paradox is that with some automated systems, the human operator is given a task which is only possible for someone who has on-line control.
- Catastrophic breaks or failures are relatively easy to identify. Automated control can, however, camouflage a system failure by controlling against the variable that is changing, so trends do not become apparent until they are beyond control.
- If a human is not involved in on-line control, he does not have detailed knowledge of current system state. The straightforward solution in the event of a detected failure is to shut down. Problems arise when, because of some factor, the process must be stabilized rather than shut down.
- It is not adequate to expect an operator to react to unfamiliar events solely by consulting the operating procedures. These cannot cover all of the possibilities, so the operator is expected to fill the gaps.
- It is ironic that the most successful automated systems, with rare need for normal intervention, may need the greatest investment in operator training.

Aviation includes a number of features that make inappropriate or failed automation more critical than most other applications. These include:

- The need for rapid action if a failure occurs near the ground.
- The inability to just shut down the system to troubleshoot and fix the problem.
- The potential for large numbers of deaths and/or injuries if an accident occurs.

The Society of Automotive Engineers (SAE) committee on Behavioral Technology has identified a number of specific concerns related to cockpit automation. The following discussion elaborates on these concerns.

Loss of Situation Awareness

Humans focus attention on the tasks they perform. They obtain information related to the task, make decisions, and take actions as a matter of course. The task of monitoring an automated system tends to be boring. If the system is reliable, only rarely is there a need for the operator to intervene and exercise his/her ability. Consequently, the operator becomes easily distracted and may tend to allocate attention to other interests. As a result, the operator may lose a sense of what is happening that is relevant to the operation for periods of time during the activity. As an operator spends months and years performing the same monitoring functions, boredom and distraction may increase, exacerbating loss of situation awareness.

Loss of Proficiency

A high degree of competence in any skill requires practice. The keen edge of finely honed skills may be rapidly lost if not used. A safe pilot needs a high degree of proficiency in psychomotor, cognitive, and communication skills. Automated systems tend to eliminate opportunities for operators to practice their skills. There is concern about how these skills will be retained in an automated system.

Reduced Job Satisfaction

A worker doesn't have to be entirely content to perform well, but satisfaction with a job is important to long-term performance and employee retention. Several factors lead to job satisfaction. These include the feeling that the job is important and that there is an opportunity of using one's

abilities to meet a challenge. Automation may have an adverse effect on job satisfaction if it reduces the opportunity to experience these feelings.

Overconfidence in the Aircraft

The negative potential of a highly reliable system is crew overconfidence in the system. The system always works, but if it doesn't, the crew can be surprised and unprepared to compensate.

Intimidation by Automation and/or Complacency

If one is below average in operating ability, deskilled through disuse, inexperienced, or overconfident in a system, he or she may be reluctant to take over when the automated system doesn't perform. If the system has been designed so that it is difficult for a flight crew to know what the automated system is doing or why it is operating in a certain way, this may add to their uncertainty. This reluctance and uncertainty will reduce the ability of the crew to fulfill its responsibility for taking over when it is appropriate.

Increased Training Requirements

Automated systems may be complex. For example, aircraft flight guidance and control systems have many modes; failures in these systems may require reprogramming or assumption of manual control. Even reprogramming to accommodate a change in the flight plan may require many actions. Thus, operators may require training in order to maintain proficiency in both the automatic and manual modes, modes which require different skills. As system complexity increases, there may be a corresponding increase in training requirements.

Inability of the Crew to Exercise Authority

This concern is related to others: the potential of automation to intimidate operators, the design of the crew interface, and fundamental design features such as envelope protection. As shown in Table 9.1, the further the crew is removed from decisionmaking by higher and higher levels of automation, the greater the danger of reducing the crew's ability to intervene or exercise authority.

Design-Induced Error

One reason designers incorporate automation is to reduce human error. Automation may reduce the frequency of human error, but the consequences

may be more critical because of the extent of the control exerted by the automation. For example, multimode displays and keyboards may require more disciplined cross checking and special procedures to assure the desired mode is selected before control actions are taken.

On the other hand, many errors attributed to humans are facilitated by poorly designed crew interfaces such as difficult-to-use displays and controls. In fact, many of the early problems that supported the development of human factors engineering as a separate design discipline were knob and dial problems. Design-induced error became recognized as a real contributor to human error. Automation does not completely eliminate this type of error, and, in some cases, may facilitate it.

Display design is one of the areas where automation may contribute to human error. The common sense approach so often used in the past by display designers will certainly not be adequate to evaluate the varieties of new format designs possible with electronic presentation. What is common sense to a designer sitting at his desk may not be common sense to a pilot flying in a crisis environment. For example, electronic displays introduced to date have presented information in formats similar to those available in conventional cockpits. These may need to be augmented by displays more suitable for the specific monitoring function required.

Developing displays with formats that facilitate quick, accurate understanding and aid problemsolving and decisionmaking can greatly enhance crew performance and acceptance of automated systems. Designing control systems that allow the crew to accurately insert information and/or control the aircraft is essential for reducing error in programming systems for normal operation as well as for making effective responses in an emergency or abnormal situation.

Design Practices

The information provided to this point indicates that although automation is advancing rapidly, it has not always lived up to its promises. One reason for this lack of complete success arises from the design processes followed. This section will consider what might be considered a typical engineering design process. While specific design teams differ in many respects, and both personnel and practices constantly change, designers historically have had

enough in common so that it is possible to identify areas where the design process can be improved, particularly during design of automated systems.

Traditional Design Approach

Historically, the main drivers of new airliner developments have been the engineering disciplines of aerodynamics, propulsion, and structures. Improved performance in speed, payload range, and efficiency have generally resulted from developments in these areas. These disciplines have therefore received the greatest emphasis (and budgets) during preliminary and advanced design phases of a project. As new customer needs are identified and technology improvements are achieved, a design team is established consisting almost entirely of designers from these disciplines.

The design team reviews accident/incident data to identify safety and reliability data that will lead to additional product improvement opportunities. They establish goals, develop alternative configurations, and perform trade studies to identify the most promising configuration(s). They calculate projected performance data and contact customers to generate or determine product interest. Customer feedback is used to refine the design. This process is iterated until a marketable design is evolved. Once enough sales have been achieved to justify the required investment and financing is obtained, the authority to proceed into the detail design and development phases is given.

The development of a new airliner may cost several billion dollars. A number of preliminary designs may be required before one is accepted for development. Consequently, preliminary and advanced design teams are usually kept as small as possible to minimize expenditure of company funds on projects that do not go forward. Frequently, cockpit human factors issues are not considered except in a cursory way during early design stages.

The average time spent from advance technical planning to certification is about 3 years. Design, subcontracting, tooling, fabrication, assembly, and testing must all be completed during this period. Generally, one year is allocated for flight testing, which reduces engineering, fabrication, and assembly to about 2 years. Thus, there is little time for research and redesign. Issues must be addressed and resolved quickly. Redesign, particularly after drawings have been released from engineering, may greatly increase costs and jeopardize contractual deadlines. For all these reasons, there is great resistance among aircraft manufacturers to changing procedures that have proven successful in prior programs.

The process described above has a number of weaknesses. One is the ability of design engineers to fully understand and weigh all the factors that influence their designs (See Table 9.8). Research into actual engineering practices has revealed a number of areas where design teams depart from the ideal (Meister, 1987). Designers often deviate from a deliberate, logical process. Behavioral data, even if available to a designer, may be ignored. Managers may reject designers' recommendations if they believe these make no difference in traditional aircraft performance parameters--reliability, cost, or development time.

Table 9.8
Cognitive Factors Influencing Design Elements (Meister, 1987)

- Statement of the problem
- Statement of criteria and priorities
- Identification of constraints
- The engineer's design style (logical, intuitive)
- Information obtained or retained
- Experience
- Preconditions (i.e., other design decisions)
- A mental outline of what must be done

Due to the revolutionary nature of cockpit changes brought about by automation and the need for experimental testing due to our incomplete understanding of human-computer interaction, cockpit design should be one of the earliest issues addressed in the design of advanced aircraft. As much time as possible should be provided to identify and address human factors issues in the cockpit. This is particularly true since human factors have received little emphasis in past designs.

Both Boeing and McDonnell-Douglas have recognized the need for increased consideration of human factors issues in the design process. These manufacturers have added to their professional staffs in the human factors disciplines and also drawn on simulation studies to support the design process.

Automation Philosophy

Past design practices have generally not made a cockpit design philosophy explicit. The general approach has been to incorporate advanced technology whenever it appeared to have a payoff or whenever the manufacturers'

customers wanted it. There has also been an interest in reducing crew workload through automation of various flight functions. The latter area received particular emphasis during the development of the MD-80 and B-757/67 designs in order to justify a two-person crew. Recently, as cockpit automation has developed and its impact on safety has generated concern, more attention has been devoted to identifying a philosophy of automation. Boeing has published a paper illustrating its philosophy for some recent aircraft (Fadden and Weener, 1984).

A primary approach to reducing flight deck workload has been to simplify system design to make the aircraft easier to operate. As an example, the number of fuel tanks has been reduced to simply fuel transfer procedures. System redundancy has been the next most common approach to increasing flight safety. Automation has been incorporated only if design goals cannot be achieved otherwise. Table 9.9 provides reasons commonly used to justifying automation in Boeing's view.

Figure 9.4 illustrates Boeing's process for determining the level of crew involvement in flight deck operations. A number of automation philosophies have been proposed for making such determinations. Table 9.10 lists some of them and their limitations.

Although none of these philosophies seems to be completely adequate at present, there appears to be growing support for the concept of human centered automation, as evidenced by the conclusions of the NASA conference attenders cited later in this discussion. It should be apparent that if an

Table 9.9
Boeing's Automation Philosophy
(Reasons to Automate)

- Simplified/minimized crew procedures for subsystem operation
  - reduces random and systematic error
  - increases time for primary pilot functions
  - prevents requiring any immediate crew action
  - reduces subsystem mismanagement accidents
  - centralizes crew alerting for error reduction
  - allows fire walling engine controls
  - allows two-person crew operation
- Improved navigation information

Table 9.9 (Cont'd)
Boeing's Automation Philosophy
(Reasons to Automate)

- provides more exact airplane position indication
- reduces fuel usage
- provides higher reliability and improves accuracy
- reduces crew error
- reduces workload, allows more preplanning

• Improved guidance and control

- reduces workload
- allows operation at lower minimums
- allows manual, semiautomatic, or automatic pilot flight
- increases precision of guidance information

effective cockpit is to be designed at minimal cost, the cockpit design philosophy should be specified early in the design process and made clear to all on the design team.

The Influence of Crew Role on Design

Design of displays and controls depends on the role that is assumed for the operator. It is therefore imperative to define the role of the flight crew operating automated aircraft prior to designing cockpit displays and controls. In designs with a low degree of automation, the operator must be present for the system to perform properly. As the degree of automation increases, the function and duties of the crew become less clear, and designers find it possible to exclude the crew from consideration. This is partially due to designers' over-confidence--their belief that systems won't fail, or that flight crews are adaptable and able to adequately resolve any problems that may arise from system malfunction or failure.

The pilot's role has traditionally been described in terms of four primary tasks: aviate, navigate, operate, and communicate. Aviate means to fly the aircraft by keeping its altitude, speed, and configuration within safe operating ranges. Navigate means to perform the actions required to fly from the present position to a desired position. Operate means to manipulate the controls required to make all of the systems--control, navigation, hydraulic, electrical, pneumatic, etc.--perform as intended and/or to compensate for equipment malfunctions.

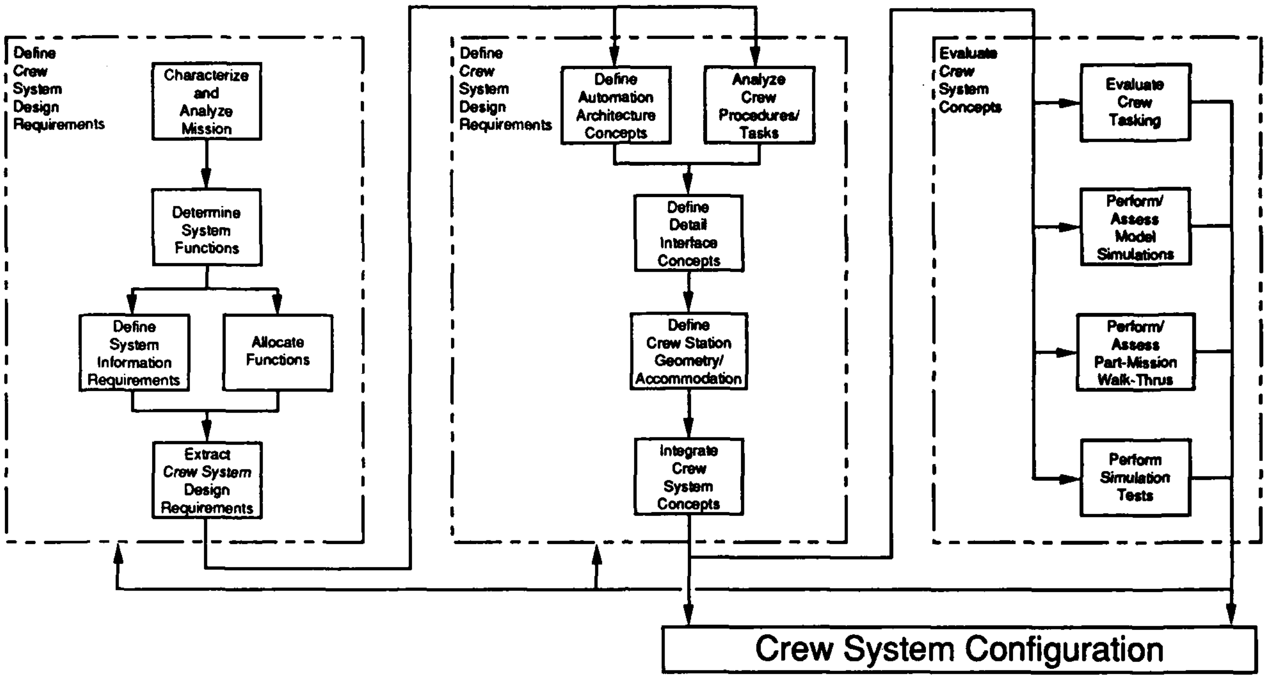


Figure 9.4. Boeing guidelines for crew function assignment.

<u>Communicate</u> means to understand human messages and interpret display information so that others inside and outside the cockpit know the aircraft's current status and intentions; it also includes providing information as required.

No system designed to date has the range of capabilities to perform all of these complex tasks. Only the human is uniquely qualified to perform the functions necessary to fly an aircraft.

Table 9.11 depicts the processes, activities, and specific behaviors that are characteristic of all task-oriented activities. An assessment of the impact of automation on the crew role reveals that flight crews are still required to perform the operator functions shown in Table 9.11. The amount and scheduling of time allocated to various tasks changes, but not the need for all of these traditional functions and activities.

A 1988 NASA conference and workshop was dedicated to identifying and addressing cockpit automation issues (Norman and Orlady, 1988). Representatives of airlines, manufacturers, pilot associations, academia, and government participated. The conclusion reached by this group was that advanced cockpits bring about both task structure and culture changes.

Some of the task changes identified were a decreased need for computations by flight crews, reduced opportunity to practice motor skills, less active systems monitoring, and more evenly balanced workload between the pilot flying (PF) and pilot not flying (PNF). In an advanced cockpit, the PF has more of a managerial function than previously, while the PNF does more work but less active systems monitoring.

Table 9.10
Design Philosophies

PHILOSOPHY	LIMITATION
Operator should be manager and make decisions at knowledge-based level (skill, rule, knowledge).	Doesn't consider operator role in compensating for system failures.
Operator should work as a manager.	Term manager is poorly defined. Cockpit management functions must play backup role in aviation since process of flying the aircraft cannot be shut down.
Let the crew do what they want to do and let automation handle the rest.	In order for concept to work, have to communicate intentions to system. Because crew desires are variable, requirements to keep computer informed may be overwhelming.
Design envelope around system. As long as crew stays within envelope, crew can fly any way it wants. If envelope is approached, computer intervenes (warns or takes control).	Variation of prior philosophy. May not be feasible from technical or cost standpoint. Envelope may vary for different routes, environments, etc.
Automate everything feasible and let crew handle the rest.	Crew may not be well adapted to assigned role. Problem of who is ultimately responsible for aircraft.
Human-centered automation.	Not defined well enough to aid designers.

Cockpit cultural changes included a more even division of responsibility, less crosschecking, and role reversal in terms of information flow, with the PNF transmitting more information to the PF than previously.

Participants in the NASA conference felt these changes were not serious in normal operations, but they might be a concern in abnormal situations involving minor and major systems failure, particularly situations involving unexpected systems failure. They concluded that it was absolutely essential for the flight crew to maintain situation dominance in all flight-related functions. In other words, the crew should have all the information and controls necessary to perform all of the traditional functions, even in automated systems. Table 9.12

presents participants' conclusions as to how the crew role has been altered in recent aircraft design.

Table 9.11
Functions of the Human Operator

PROCESSES	ACTIVITIES	SPECIFIC BEHAVIORS	
PERCEPTUAL	SEARCHING FOR AND RECEIVING INFORMATION	DETECTS INSPECTS OBSERVES READS	RECEIVES SCANS SURVEYS
	IDENTIFYING OBJECTS, ACTIONS, EVENTS	DISCRIMINATES IDENTIFIES	LOCATES
	INFORMATION PROCESSING	CATEGORIZES CALCULATES CODES COMPUTES	INTERPOLATES ITEMIZES TABULATES TRANSLATES
MEDIATIONAL	PROBLEMSOLVING AND DECISIONMAKING	ANALYZES CALCULATES CHOOSES COMPARES	COMPUTES ESTIMATES PLANS
COMMUNICATION		ADVISES ANSWERS COMMUNICATES DIRECTS INDICATES	INFORMS INSTRUCTS REQUESTS TRANSMITS
MOTOR	COMPLEX-CONTINUOUS	ADJUSTS ALIGNS REGULATES	SYNCHRONIZES TRACKS
	SIMPLE-DISCRETE	ACTIVATES CLOSES CONNECTS DISCONNECTS	JOINS MOVES PRESSES SETS

Human Factors

Human factors may be defined as the application of knowledge about human characteristics to the design, operation, and maintenance of systems. This discipline gained recognition during World War II when the military recognized that performance and safety could be enhanced by improving the harmony between machine and human characteristics.

Initial human factors interest was largely in knobs and dials--in improving displays, such as altimeters, and controls such as levers, knobs, and cranks. Fundamental principles such as control-display compatibility, and color and position coding, are products of this work. Many of the early contributors to human factors were experimental psychologists drawn from academia and employed by the armed forces to study specific problems. At the end of the war, most of these professionals returned to civilian status. A few remained in government laboratories.

Table 9.12
Crew Role

	HISTORICALLY	WITH AUTOMATED AIRCRAFT
PRIMARY RESPONSIBILITY	SAFETY	SAME
PRIMARY FUNCTIONS	AVIATE NAVIGATE COMMUNICATE OPERATE	SAME SAME SAME SAME
PRIMARY TASK CHARACTERISTICS	DIRECT CONTROL	INDIRECT CONTROL
	MANAGER, OPERATOR	MANAGER, MONITOR
	PRIMARY BACKUP TO SYSTEMS	SECONDARY BACKUP TO SYSTEMS
	DIRECT INVOLVEMENT CONTINUOUSLY	INTERMITTENT DIRECT INVOLVEMENT
	MULTIPLE SOURCES OF INFORMATION	FEWER INFORMATION SOURCES
	INFORMATION GENERALLY AVAILABLE	INFORMATION MAY HAVE TO BE RETRIEVED
	PERCEPTUAL/PSYCHOMOTOR SKILLS USED FREQUENTLY	PERCEPTUAL/PSYCHOMOTOR SKILLS NOT DEMANDED VERY FREQUENTLY
	CAPTAIN'S AUTHORITY FINAL	CAPTAIN'S AUTHORITY MAY BE PARTIALLY ABROGATED

As system complexity increased, the U.S. Air Force recognized the need for more emphasis in human factors and mandated contractors to employ specialists in this area. Few schools offered courses in the discipline and companies found it difficult to employ properly qualified people. There was uncertainty also regarding the role and organizational placement of human factors specialists. Often they became internal consultants who were used to make recommendations or perform studies to solve problems after these were identified.

Because solutions to these problems called for consideration of many issues and an adequate database was not available, the human factors specialists

recommended experimental studies to resolve the issues. Contributing to their desire to experiment was the fact that most were trained as scientists, not as engineers. Design engineers, on the other hand, often couldn't afford the time and/or budget to accommodate experiments. In addition, human factors specialists often did not have a background in either aviation or design. As a result of these and other considerations, the human factors discipline has been slow to gain whole-hearted support from either the design or the operational communities.

As system complexity has continued to increase, however, the need for consideration of human capabilities and limitations has been increasingly recognized. Most large aircraft manufacturers now maintain a human factors staff. In fact, human factors disciplines have expanded to include not only experimental-industrial psychologists, but also physiologists, anthropometrists and other life and social scientists. Some human factors departments also include aerospace medicine physicians and training specialists because of the commonality of their interests and academic backgrounds.

Human factors staffs have become much more knowledgeable about flight operations and design constraints as they have grown in experience. As a result of their increased knowledge in these areas, they have also become more responsive to management and design needs.

In spite of these advances, however, most organizations do not accept human factors as a core discipline with a status comparable to structures, aerodynamics, avionics, and more traditional engineering disciplines. (One exception to this is the U.S. Air Force which has established the Human Systems Division as one of its prime Research and Development organizations.) Organizations often support human factors only reluctantly. There are many reasons for this reluctance:

- Overconfidence in the human ability to adapt to their designs
- Faith that training can compensate for design shortcomings
- Belief that human factors involves only common sense
- Belief that the sciences upon which human factors is based are "soft" and pilot experience is better than human factors data
- Judgment that the system will benefit more from an additional engineer from one of the traditional engineering disciplines than from a human factors specialist.

Perhaps the most significant contributor to organizations' reluctance to support human factors is the lack of objective human factors criteria in the Federal Aviation Regulations or in typical design specifications.

There is ample evidence that attention to human factors is warranted based on accident and laboratory (including simulator) data. It would seem prudent, for example, that designers should invest most heavily in the area which has been found to be the largest contributor to accidents, i.e., human error. Since many of these errors result from design-induced causes, human factors should be a major concern of both designers and certifiers.

How Human Factors Relate to Automation Design

Flexibility and adaptability are prominent characteristics that make humans essential to a system. But this flexibility and adaptability are achieved at a cost. There are design trade-offs for humans as there are for other systems components. One of the reasons humans are so flexible and adaptable is their complexity. Many interacting variables can influence their behavior and performance. This section will briefly address some of the most fundamental human characteristics related to working with automated systems.

In many automated systems, the role of the human is that of a monitor. If something fails, the human is expected to detect the failure, determine the problem, decide what action to take, and execute the appropriate response. The human must act as a sensor, decisionmaker, and controller. The performance of humans as monitors has been studied extensively since World War II. The development of radar and sonar put some people in a position where it was necessary to detect small stimulus changes which do not occur very often. Experiments investigating human performance in this type of situation became known as vigilance studies. Generally, it has been found that humans do not perform well as monitors. If their interest is not maintained, they become easily bored or distracted and direct their attention to other considerations.

Physiological studies have determined that periods of inactivity with few demands are not conducive to good performance. The need for stimulus change may cause the monitor to attend to other nonwork-related interests. Attention to an outside stimulus may make it difficult physiologically for work-related stimuli to be perceived. The brain inhibits the neural response to stimuli that are not related to its primary focus (Hilgard and Atkinson, 1967). There is also evidence that if attention is dedicated to one channel of information for a period of time, information from other channels may tend to receive increased priorities (Broadbent, 1957).

An additional human cognitive characteristic is the need for warm-up. Once a task has been deferred for a while and then reinitiated, it takes some time before the person is able to perform the task at peak effectiveness. The degree of cognitive warm-up required depends on the person's level of skill, the difficulty of the task, the time since performing the task, the degree of similarity between the task and intervening activities, and other factors.

These and other findings lead to the conclusion that too low a workload degrades human performance. Similarly, if the workload is too high, performance can suffer. Optimal performance generally is obtained when the relevant variable is in the middle range. Figure 9.5 illustrates this relationship for workload. Figure 9.6 is an example of how the relationship could be applied to cockpit design.

Hypothetical Relationship Between Workload and Performance

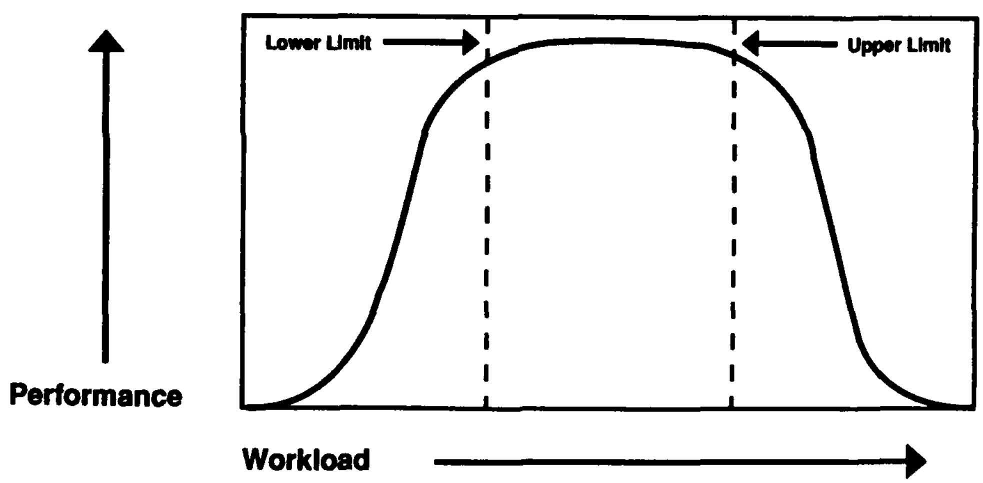


Figure 9.5 Hypothetical relationship between workload and performance. (original figure)

Research dedicated to finding valid, reliable measures of workload has been recently emphasized in a number of laboratories and progress has been made. A recent FAA contract supported an extensive review of the workload measurement literature (Corwin, et al, 1989). Physiological, behavioral and task analysis measures were investigated. Although no totally acceptable assessment method was identified, a number of useful techniques are available.

Table 9.13 is a list of psychological phenomena relevant to human performance with automated systems. The literature includes a great deal of research in each area. Understanding and interpreting this literature requires specialists. This need for specialists is becoming increasingly recognized by many agencies. The

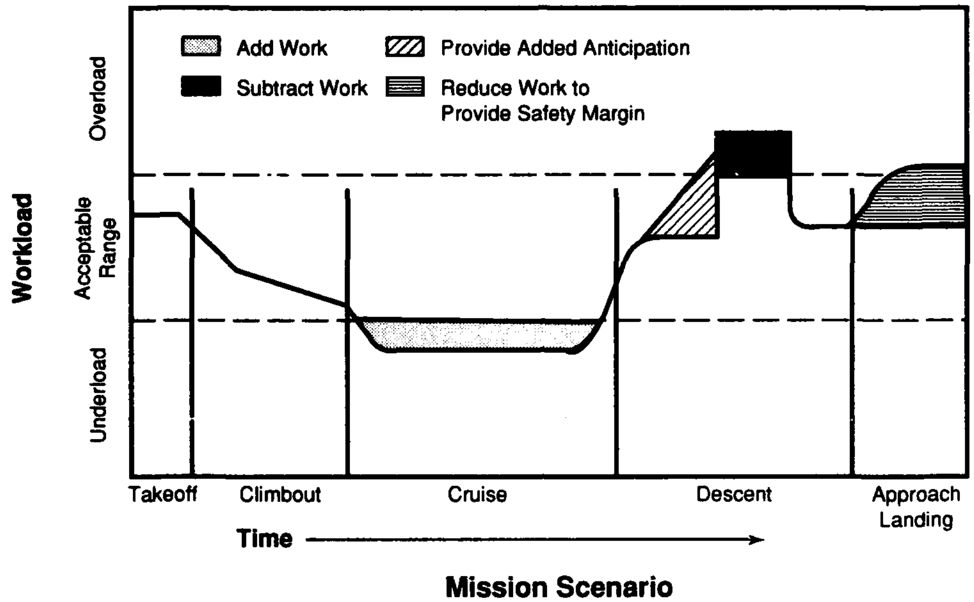


Figure 9.6 Beneficial adjustment of flight crew workload by phase of flight. (original figure)

Air Transport Association, for example, has not only established a standing task force to identify human factors issues and promote their resolution, it has encouraged the elevation of human factors to a core discipline in aircraft design commensurate to such engineering disciplines as aerodynamics.

"Soft" Sciences and the Need for Testing

One of the reservations many organizations have about human factors is that they are supposedly based on "soft" sciences. This perception is not accurate. Research into human characteristics has generated a great deal of "hard" information. Sensory processes are reasonably well understood, and a great deal is known about perception, learning, memory, motivation and emotion. Useful data are also available regarding decisionmaking.

It is true, however, that in spite of the amount of data available, few theories are available to integrate these data in useful human factors applications. Adding to the difficulty is the fact that many interacting variables may influence a person's performance in unpredictable ways at any specific time. These

limitations make prediction of an individual's behavior at any instant difficult. This problem is not unique to the behavioral and social sciences, however. Medicine and pharmacology are similarly affected. More to the point, perhaps,

Table 9.13
Some Psychological Topics Relevant to Automation

- Arousal
- Motivation (Yerkes-Dodson-Law)
- Stress
- Inhibition (Hernandez De Peon)
- Inverted u-shaped curve
- Attention
- Isolation
- Vigilance
- Overload
- Sensation and perception
- Influence of learning/practice on perception
- Response time and mental set (anticipation)
- Warm-up
- Short-term memory and distractions
- Long-term memory
- Need for practice
- Transfer of training
- Stress
- Biases in decisionmaking

is the fact that many "hard" disciplines such as aerodynamics and meteorology also have similar problems. In all of these disciplines, there is a need for extensive testing to determine the efficacy of a particular design or model.

Testing is heavily emphasized in most aircraft design. Aircraft structure is stressed to destruction at a cost of many million dollars to demonstrate that design requirements are met. Millions of dollars a week are spent on wind-tunnel testing during some phases of design. In contrast, simulator tests of cockpit design have not been as frequently or effectively used as other modes of aircraft design testing. This seems inconsistent in view of the much greater confidence in the "hard" data of the more traditional disciplines and the identification of human error as a major contributor to accidents.

The Problem of Criteria

Frequently, nonspecific criteria are used in making human factors assessments. Subjective pilot judgments are probably the most frequently used criteria for design acceptability. Pilot judgment has a number of advantages from an engineer's point of view, such as its apparent validity and quick response. Rarely, however, are project pilots knowledgeable in the scientific disciplines of experimental psychology, physiology, and anthropometry that are the foundation of the human factors specialists. Test pilots, while well trained for their job, may lack an understanding of the line pilot's environment, such as flying the same aircraft for years, flying many legs late at night, or flying long intercontinental flights.

Ideally, design decisions should be based on criteria related to overall system performance, but designers have generally deemed human performance difficult to assess. Part of the difficulty may arise from the designers' relatively poor understanding of human factors testing. It seems apparent that more attention should be devoted to valid, reliable human performance measures.

In addition, if critical human tasks involve reprogramming and/or taking over for automated systems in the event of a significant failure, it should always be demonstrated that representative crews can perform adequately under representative (including worst-case) scenarios. It is also desirable that human performance be tested near the limits of its capabilities to assure adequate safety margins.

Conclusions

This review far from exhausts the relevant information regarding cockpit automation. Training issues have not been addressed at all, for example. Many years of further study and of industry experience will be required for designers to be fully confident in how to design automated systems that are compatible with human characteristics. Several preliminary conclusions seem appropriate, however:

- Automation will continue to increase.
- Successful automation depends on proper integration of human capabilities.
- The discipline of human factors has a store of knowledge and methods which can be useful to good systems design.

- Automation is generating and/or highlighting human factors issues.
- Attention to human factors has generally been lagging in both design and certification.
- Adequate emphasis on human factors will be facilitated by a System Engineering Approach, established human factors performance criteria, early investment in cockpit definition and development, developmental simulation and testing of human factors issues, and the development of a design-oriented Human Factors Research and Development program.

Recommendations

This review suggests several basic automation and human factors questions that need to be addressed in certification:

- Will the crew be exposed to potentially catastrophic failures in which their actions will be crucial?
- If so, will the crew be able to execute the appropriate actions in the time available without making a catastrophic error?
- What is the probability of each of the above?
- Based on the responses to these questions, will safety be degraded or enhanced by the automation?

To obtain valid, reliable answers to these questions it is necessary to consider not only the aircraft features but the whole aviation system. This consideration should include crew functions in normal and abnormal operations; interactions between the crew and the system; and crew selection, training, and composition. Without question, the FAA certification process provides for this. However, the process needs to be strengthened in several areas if automation and human factors issues are to be adequately addressed.

First, consideration should be given to making the human factors certification criteria more explicit and objective. These criteria should be stated in terms of performance rather than rules of design so as to minimize difficulties as technology advances.

Second, the FAA should add human factors specialists to its certification team. The basic issues supporting this recommendation have been reviewed in this

discussion. Attention to human factors by the Agency would encourage manufacturers to increase their emphasis in this area, and the addition of basic human factors knowledge would greatly enrich the FAA's assessment process. Table 9.14 identifies activities the FAA certification specialists could perform.

The FAA has recently added a human factors specialist at Headquarters. Although human factors emphasis on policy and research is desirable, the real payoff will be obtained only if human factors are incorporated in the certification process.

Table 9.14
Role of FAA Cockpit Certification Specialist(s) - Human Factors

- Review, critique, assess, and enrich manufacturer's cockpit development plan
- Participate in selected development activities to assure adequacy
- Review and assess cockpit relevant reports of tests, analyses, etc, submitted by manufacturer
- Participate in development of FAA cockpit certification requirements
- Participate in certification testing:

  -experimental design

  -definition of criteria and performance measures

  -adequacy of statistical analysis

  -subjective assessment methods

  -human factors checklist(s) development and application
- Noncockpit certification activities:

  -helping to identify and structure FAA/NASA human factors R&D efforts

  -identification of crew training issues
- Alternative approach - DERs for human factors

Chapter 10

Display Design

by Delmar M. Fadden, Chief Engineer--Flight Deck, Boeing Commercial Airplane Group

The rapid and reliable display of visual information in the flight deck requires a thorough understanding of the functions being supported, thoughtful application of available human performance knowledge, and careful selection of the appropriate display media. This chapter explores some display characteristics of special relevance to achieving highly effective human performance in flight situations.

The measure of a truly effective display is how well it supports consistent accomplishment of the tasks assigned to the person who will be using it. Display design is as concerned with task design as it is with presentation symbology and display devices. The process of identifying the full range of tasks and the associated information requirements for a modern, highly integrated display can be formidable indeed.

The core of the design process usually involves resolving contentions between system functional requirements and operator capabilities and limitations. Through actual design examples, this chapter illustrates the issues associated with balancing system and human needs. The examples are based on display development work at Boeing Commercial Airplane Group in support of the 757, 767, and 747-400 airplanes.

Display Development Process

Figure 10.1 is a flowchart representing the primary display development activities (ARP 4155, SAE G-10 Committee, 1990). The flowchart provides a useful basis for discussing the fundamental elements of display design. Some steps require considerably more effort than others, depending largely on the scope and phase of the project. Some of the steps can be accomplished using traditional engineering tools and methods; others are better suited to techniques more commonly associated with the sciences of psychology, operations research, and human factors. Many successful displays have been developed without explicit attention to this process, though their development histories often show evolutionary improvements that can be mapped to these steps.

Requirements

Displays exist to provide information to a human being who is asked to achieve some objective. Accurately recognizing that objective in terms of required outcomes is crucial to successful display design. Once the top level objectives are identified, the focus shifts to determination of the detailed tasks necessary to accomplish the objective and the information requirements that support those tasks.

There is an understandable tendency to skip the formal definition of the detailed tasks and associated information requirements and start the design by developing display formats. Working on display formatting can be a useful aid in initiating an understanding of the information requirements. However, the understanding gained by first developing information requirements from the related tasks is virtually always more accurate and complete. The are two significant side effects that can follow a design which begins with display formatting selection. The format selected likely will be based on the similarity of information content to that of other displays rather than any actual linkage to the tasks this specific display supports. The conceptualizations of the information required and its organization will be well established before the full range of task possibilities has been explored. Together these effects can result in excessive display complexity, more operator errors, and less efficient task performance by the pilot.

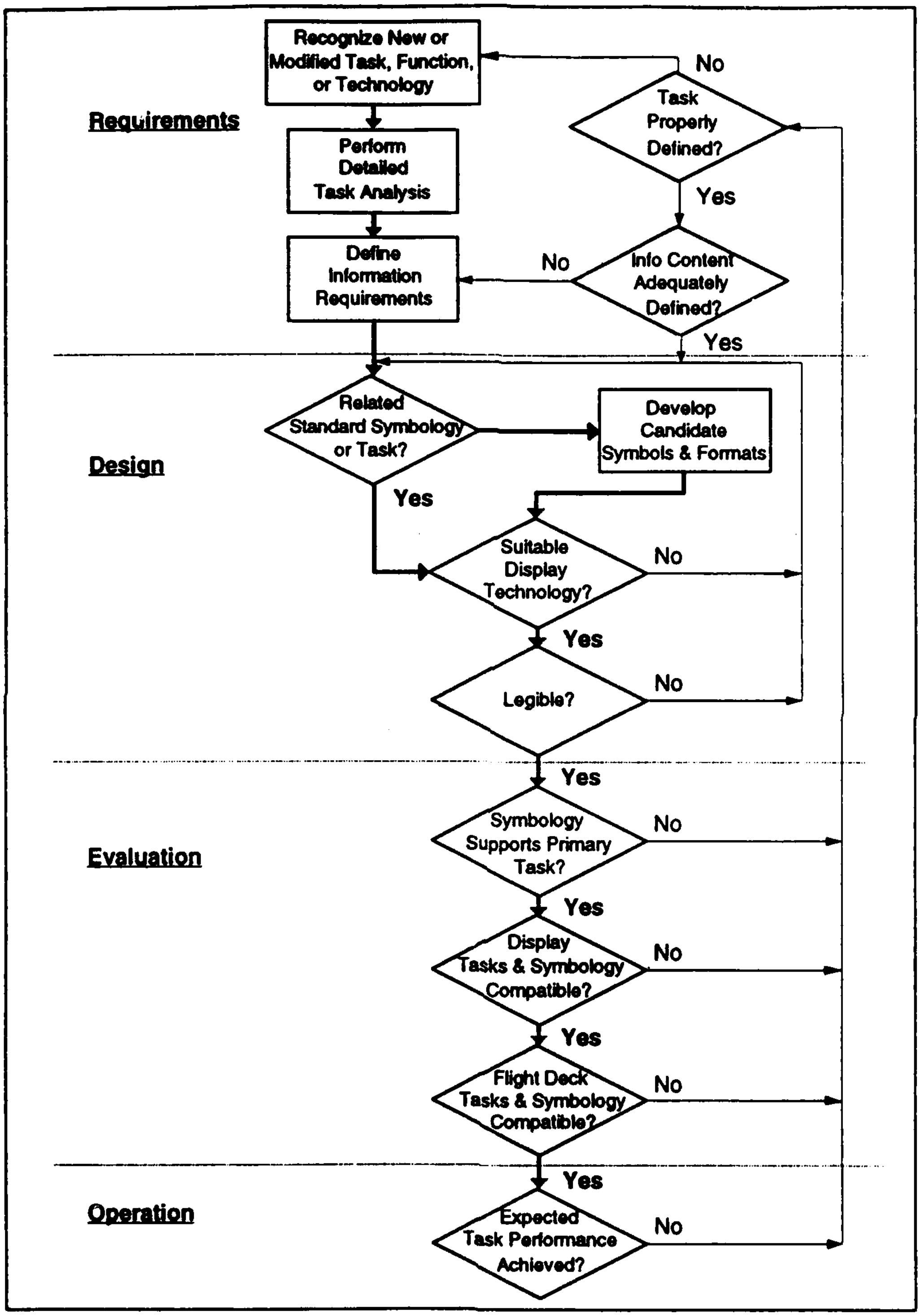


Figure 10.1 Display Development Flowchart. (from SAE document ARP 4155).

Each element of information to be displayed is integrally linked to the dynamic performance requirements for the associated task. Characterization of the task with measurable performance objectives is a key step in understanding the specific contribution of each information element and the effectiveness of the symbology used to portray that information.

In most circumstances it is best to trace the task from the top level mission objectives. While this is tedious work, it provides essential insight into the interrelationships between tasks and provides a basis for meaningful discussions with those who will use the system. The top-down task analysis also yields a complete description of the steps believed necessary to execute each particular task. To some, it may seem premature to prepare a task analysis before the hardware is designed. However, the system designer knows conceptually what has to take place. Specific switches and controls are not yet known, neither are the overhead tasks which will be necessary to operate the system, so the initial analysis must be done at a top level. As the design develops, the analysis can be expanded to a more detailed level. Proceeding in this fashion provides a good check on the correctness and efficiency of the specific design. By comparing the detailed analysis with the initial top level analysis, the designer determines how much overhead has been added and checks to see that the functional design remains consistent with the stated objectives.

The task analysis should identify at least the following information:

- the objective of the task, stated in measurable terms;
- the timing for the task (any task initiation dependencies should be defined, along with constraints on execution or completion time);
- expectations about task performance, including accuracy, consistency and completeness;
- possible errors and related consequences (be sure to consider errors of omission along with errors of commission);
- task dependencies, other than those associated with timing already identified (dependencies might include other tasks, specific events, combinations of flight conditions, etc.);
- criticality of the task objective in relationship to the safety and efficiency of the flight.

Having completed a detailed task analysis, the tasks can be linked with the information necessary to accomplish each task. This would normally include:

- o definition of information requirements
- o the accuracy and range needed
- o the context within which that information will be used
- o the necessary dynamic response
- o any special relationships with other events or information

At this point, it is useful to examine similar tasks and the information necessary to support them. Task similarities provide valuable insight into the range of human performance that can be expected. Where tasks are new or involve a change in the required precision or dynamics, the designer will have to turn to rapid prototyping, part task simulation, experimental tests, or other human factors testing to identify and quantify the specific information requirements. Tasks involving continuous dynamic control are further complicated by the complex interaction between the *dynamics of the control* device, the vehicle dynamics, the dynamics of the displayed information, and sometimes the dynamics of the pilots' response. In difficult cases, this step will be iterated many times in a series of progressive refinements until satisfactory performance is achieved. Often these iterations are accomplished in conjunction with iterations of the previously discussed task analysis and the symbology development step that follows.

Not all information requirements need to be satisfied through on-board displays. There are various other sources for required information that can be just as effective. One of these sources is the knowledge that the pilot carries in his or her mind through previous experience or training. Also, information can be carried on board with the pilot or the pilot can derive it from other information available on the flight deck. Information from an alternate source may be easier for the pilot to integrate with the task than if it were contained in a flight deck display. Taking the time to examine alternate sources for required information can simplify display design considerably and aid the pilot by simplifying access to the required information.

Design

Once the information requirements for a display have been defined, the next step is to determine how to present the information. Symbology selection determines how specific information elements will be represented within the display. (In this context, symbology encompasses any form of character, graphic, or textual entity.) By contrast, display format selection determines the conceptual framework within which the information will be presented. The two selections are closely related. For highly integrated displays, the selection of formatting will be heavily influenced by the top level tasks the display supports, while the symbology selection often will be guided by specific requirements of the detailed tasks. The necessity for joint and iterative refinement of symbology and formatting frequently increases as display complexity increases.

It is standard practice to pay particular attention to how information has been represented and related to tasks in similar successful displays. Building on past successes has numerous advantages. Training can be simplified, if the pilot is familiar with a significant portion of the display. The risks of introducing a new display can be reduced, if the human performance expectations are based on operational use of a similar display. These benefits are often perceived to be of sufficient value as to preclude serious consideration of alternative symbology and formats. However, examine the underlying tasks carefully, since subtle differences in the current tasks may require that different information be portrayed or that formatting be adjusted to highlight different relationships. Changes in the technology used for display can force a change in the selection of symbology even when the task and information requirements remain the same. This would be the case when the change in technology alters important characteristics used in creating symbology. For example, line widths that can be presented using practical CRT technology are considerably thicker than those which can be produced using print technology. This changes the amount of detail that can be presented successfully in a given area. In effect, print media have a much greater upper limit for information density when compared with a CRT display. Another difference concerns the manner in which displays generate brightness. Since CRTs emit light, the overall brightness of a CRT display will be a direct function of the information content and a reverse function of the ambient light. Reflective displays, on the other hand, change brightness as a direct function of ambient light with a much smaller contribution based on the information content.

When technology changes are involved, it should not be assumed that symbology that has been successful in the past will carry over equally successfully to the new display. Each display technology has unique characteristics or capabilities that can be exploited to enhance the effectiveness

of information transfer. The common ground for assessing the impact of any limitations and the value of any enhancements is the task performance achievable. Objective evaluation of these issues has a profound impact on the decision about which technology to use.

If there isn't an existing presentation format for the task, new symbols and formats must be created. Simplicity, quick recognition, and directness are characteristics of proven value in effective symbology. Regardless of how the symbol is conceived, there needs to be an appropriate performance measure (agreed to in advance) to determine how well the symbol performs its job. User preference is a significant factor in the development of symbology. If the users don't like a symbol, there is little to be gained by continuing its use. However, just because the users like a symbol does not mean that they can use it effectively. The only way to know that a symbol really works is to have the pilot use it and to measure the resulting performance.

As in all human performance testing, the test engineer is faced with the challenge of obtaining an appropriate performance measurement yardstick. In this case, it comes from the detailed task analysis. How much tracking accuracy does the pilot have to achieve? What probability of error can be tolerated? How quickly do decisions have to be made? These questions can be quantified based on the pilot's top level task and the details of the task analysis.

Finally, designers have to look at factors of legibility, so that the displayed information can be seen in the operating environment. Legibility is a complex issue in a modern airplane. Several factors contribute to the potential for less than optimal viewing: the geometrical requirements for the aerodynamic shape of the flight deck, external environmental influences, and the large vision variability between pilots. Vision is one of the more variable of human capabilities. It is not unusual for otherwise similar pilots to have quite different visual capability. Corrective lenses can reduce the effects of individual acuity differences; however, accommodation time, color perception, and critical flicker fusion frequency remain highly variable individual characteristics. The pilot's external environment varies from virtually pitch black to extremely bright sunlight. The distance between the pilot's eyes and the display is generally greater than a person would choose to read a book or a newspaper. Accordingly the size of text and graphics must be increased to compensate. The pilot and the displays vibrate at different rates when the airplane is in turbulence. The resulting relative motion can severely hamper readability, particularly the readability of small symbols or fine detail. If all or a portion of the information must be read in turbulence, both the design and the legibility testing must take that into account.

Evaluation

The first part of the evaluation cycle determines whether the primary task performance defined in the early requirements phase has been achieved. As with the early development work, it is having clearly identified and measurable performance criteria, that makes efficient testing possible. Once it has been determined that the expected performance can be achieved for the intended task, it is important to determine that the performance of other tasks has not been degraded. This second portion of the evaluation process is generally more difficult.

Knowledge of the various mechanisms that have contributed to performance degradation in the past is a good place to start in developing an evaluation strategy. Typical conflict mechanisms include the following:

- Apparent symbol motion caused by actual motion of nearby symbols.
- Poor recognition of a symbol or alphanumeric caused by excessive dominance of an unrelated nearby symbol. Such dominance may be due to relative size, color, brightness, or shape differences.
- Symbol uses or format interpretations that are inconsistent with pilot expectations. The pilot's expectation derive from many sources including: other associated tasks or displays, his mental conceptualization of the situation, cultural influences, training, or previous experiences.
- Similar symbols that support different tasks but can be confused. This problem is particularly difficult to identify if the information is identical or highly similar but the task or task performance level is subtly different.

Integrated displays present a great deal of information, and have many tasks associated with them. Therefore, if a new task is being added to an already complex display, it is important to confirm that the required level of performance for previous tasks can still be achieved. Once task performance has been confirmed for all tasks associated with the integrated display, the check should be expanded to examine all applicable task-display combinations on the flight deck.

Operation

The final phase in the display development process is operational follow-up. Comments about problems or concerns are readily available from both certification and airline personnel. Over the life of a typical display system, it is not unusual to find that some of the tasks for which the display was originally designed get redefined in subtle ways. This may be due to changes in the operating environment, changes in the skill or knowledge base of the pilots using the display, or it may be the result of refinement of a partially understood task. In any case it is important that user comments be recognized and evaluated against the design intent. The quality of future decisions about use of the display, associated training, or operational enhancements depends on accurate understanding of the pilots' tasks and how they are supported by displayed information.

General Design Issues

Opportunities for Standardization

A recurring question directed to display designers deals with the notion of standardization. A typical question might be, "Since these displays contain the same information, can't we standardize on common *symbols and formats?*" The answer focuses on the detailed nature of the pilots' tasks. If the tasks are indeed common, then a common display would be operationally attractive. However, if the tasks are different in any significant way, a standardized display may result in degraded pilot performance or an increased error rate.

An example may clarify this counter-intuitive situation. The relative merits of vertical tape engine instruments as compared with round dial displays have been debated by developers since the mid 1960's. With the appearance of CRT engine displays in the 1980's, it became feasible to provide whichever format an airline preferred. During development of the 757 and 767, Boeing conducted numerous simulator tests and demonstrations designed to aid airline personnel in the selection of the preferred format. The unanimous selection was the round dial format, in spite of initial pilot expectations that a more balanced preference would exist. Similar simulation testing was conducted during development of the 747-400 in the late 1980's. This time, the unanimous selection was the vertical tape format.

The engines on the 767 and 747-400 are identical; the same part number. The parameters displayed to the pilot are identical. Why should there be such a marked difference in selection? As discussions between pilots and researchers probed this issue, it became clear that while the high level task objective is

indeed the same for the engine-related tasks on the two airplanes, the task execution strategies the pilots preferred were distinctly different.

For the four-engine 747-400, the pilot monitored for an engine anomaly by comparing the same parameter on all four engines and focusing on the engine whose parameter was inconsistent with the other three. For the twin-engine 767, the strategy involved comparing the parameters for each engine with the pilot's expectations and his knowledge of past performance. In this case, the pilot was concerned with relating the different parameters for a single engine. Cross comparisons for the twin-engine airplane would be inconclusive for many failure conditions.

Understanding this difference in task execution strategy provides a good basis for understanding why there was such a clear difference in the display format selection for the two airplanes. Where task differences do exist, the issue of standardization can be reduced to comparing the cost saving which might result from standard display hardware and software with the cost of the associated degradation in performance and the added compensatory training that would be necessary.

Flight functions that are common across many airplane types come under significant market forces that, over time, promote de facto standardization. This tends to apply to functions that are well known and quite stable. As would be expected, the bulk of industry attention is focused on functions that are new, incompletely understood, and rapidly changing. It should be possible to achieve a reasonably high level of display standardization provided that detailed tasks can be standardized. The crucial factor is whether the tasks are truly common. That is a difficult question to answer in a business climate involving intense competition and rapid technological change both on the flight deck and in the ATC environ. In many ways, it is a tribute to the entire industry that the degree of standardization that exists now has been achieved at all.

An example illustrates the subtlety of the pilot's use of dynamic symbology. The primary instrument arrangement for the Boeing 767 has the map display directly below the primary attitude display. The localizer deviation display is at the bottom of the ADI. Since the track scale is at the top of the map display, there is no need for repeating any heading information on the ADI. The Boeing 747-400 has larger CRT displays in a side-by-side arrangement. In this case, the track scale is separated from the localizer deviation. Since this altered the "basic T" instrument arrangement, it was decided to place a heading scale at the bottom of the primary flight display (PFD). The initial format for this information was selected to emphasize airplane heading, thus maintaining a strong link with past HSI displays. The scale at the top of the navigation

display (ND) is track oriented, as it is on most 767 airplanes. The two different orientations were believed to match a difference between the localizer capture and runway alignment tasks. In separate applications, both of these orientations had been in wide-spread use for an extended period of time, each with highly successful operational histories. During initial 747-400 flight testing, it was found that a significant number of pilots were having difficulty with the transition between instrument and visual conditions during initial departure and the final phase of ILS approaches. Having the two scales in close proximity and with a different orientation was suspected as contributing to the problem, since the basic information contents of the displays on the 767 and the 747-400 are identical. Identification of the specific sources of the performance difficulty was done by a team led by John Wiedemann. The steps they accomplished in resolving the difficulty provide an interesting perspective on the complexity of designing highly integrated displays.

Figure 10.2 shows the original 747-400 heading and track symbology on the primary flight display (left side of the figure) and navigation display (right side

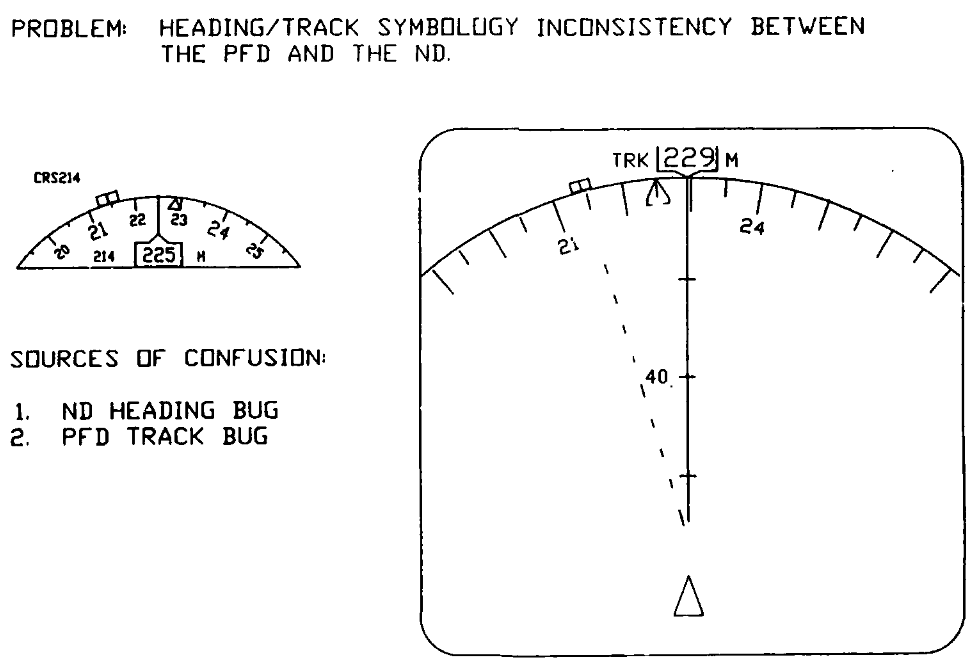


Figure 10.2 Initial 747-400 PFD and ND Heading and Track Symbology. (original figure)

of the figure). On the navigation display, track is fixed at the top of the display and heading is shown by a modified, triangular pointer which moves along the

inside of the compass scale. Conversely, on the PFD heading is indicated by a fixed pointer (which appears as a mirror image of the ND track pointer) and track is shown by a small triangular symbol which also moves along the inside of the compass scale. In both displays, the selected heading value is indicated by a split rectangle which moves along the outside of the compass scale. On the ND, the selected heading is reinforced by a dashed line emanating from the airplane position and leading to the split rectangle.

The first step in clearing up the problem was to minimize confusion caused by the different pointers by changing the ND heading pointer to make it more distinctive. Figure 10.3 shows the new shape.

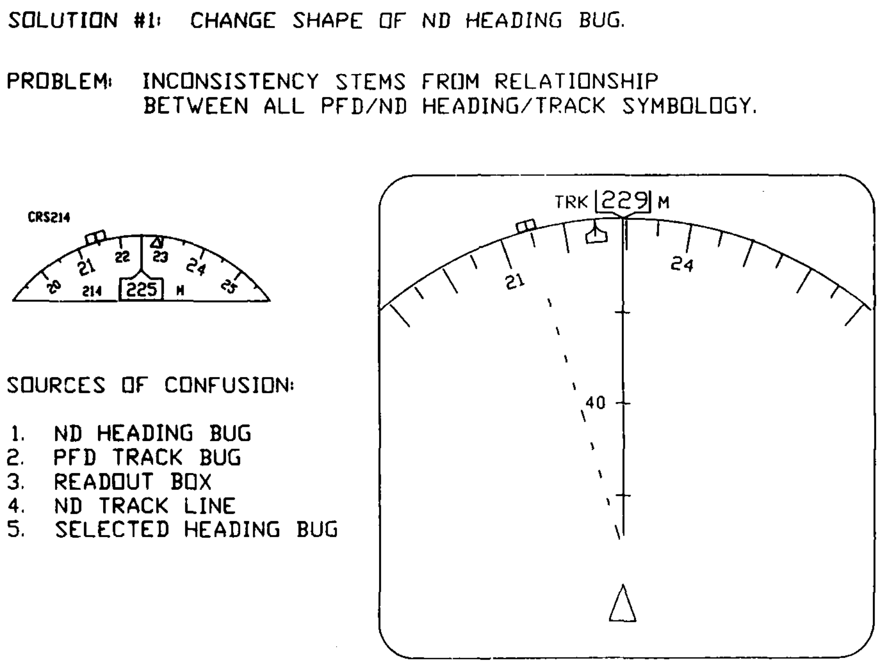


Figure 10.3 Navigation Display Heading Pointer Shape Change. (original figure)

This simple change did not solve the problem. At this point, a thorough review of task information relationships was accomplished beginning with an assessment of how these tasks were supported by earlier displays. This review confirmed that the information content was correct but indicated three areas of potential confusion brought about by the close proximity of the PFD and ND presentations. The next step involved changes in each of the three areas: (shown in figure 10.4)

- o make both heading pointers the same shape, but put them outside the compass scale circle,

- o locate both digital readout boxes at the top of the display,

- o add a moveable track line on the PFD, analogous to the fixed track line on the ND.

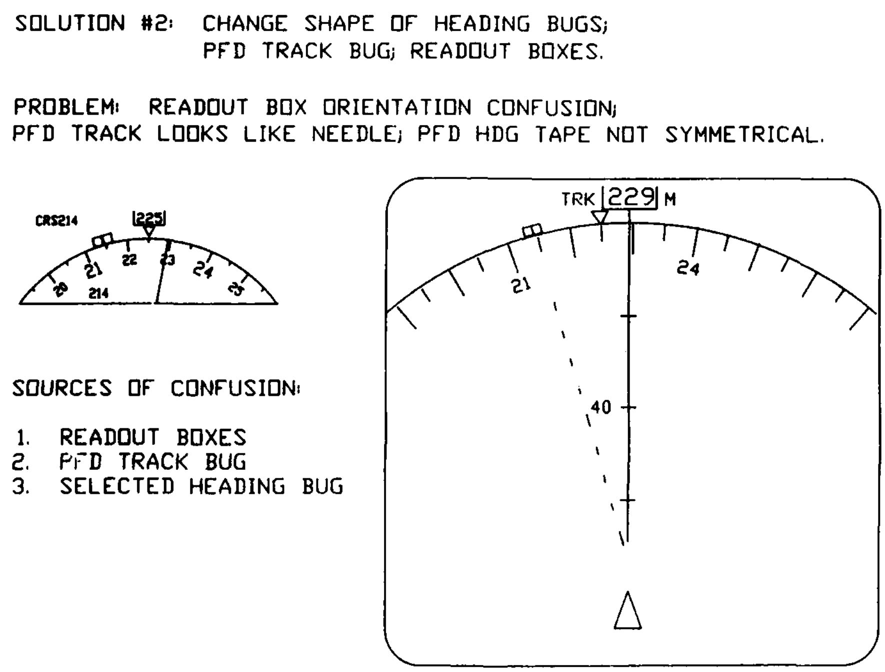


Figure 10.4 Consistent Shapes for Heading and Track Pointers. (original figure)

Performance with this format was better; however there was now confusion associated with the digital readouts and the track information. The results of simulator testing suggested three more changes: (shown in Figure 10.5)

- o remove the digital readout box from the PFD, so there is no read-out confusion;

- o add a tick to the PFD track line to strengthen the association with the ND track line;

- move the selected heading split rectangle to the inside of the compass arc to avoid conflict with the heading triangle.

FINAL SOLUTION: CONSISTENT SYMBOLOGY BETWEEN DISPLAYS

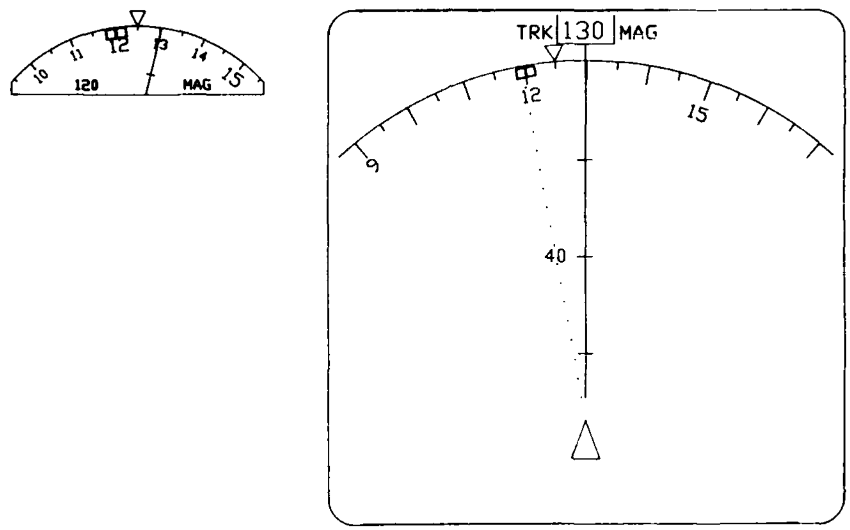


Figure 10.5 Consistent 747-400 PFD and ND Heading and Track Symbology. (original figure)

This combination performed well. Clearly the success of this symbology suggests that the actual task for which the pilots use the scale on the PFD is closer to the capture and track task associated with the map display than the runway alignment task that had been presumed. This interpretation follows from the relatively small change in the ND format compared with previous map displays and the much more significant changes to the PFD when compared with previous HSI presentations. Note that no information was added or removed from either display. All seven of the changes involved symbology and formatting only. The number of changes and the sequential manner in which they were identified emphasizes the high degree of interaction among the symbols in these two displays.

Use of Color

The first commercial CRT displays developed by Boeing (originally intended for the Boeing SST) were integrated in the NASA TCV airplane after the SST

program was canceled in the early 1970s. (The NASA Terminal Configured Vehicle, TCV, is a Boeing 737 airplane with a reconfigureable research flight deck and extensive avionics designed to address a wide range of systems integration and pilot performance issues.) The TCV CRT displays were monochromatic because there was no suitable color display on the market that could be viewed in even bright room light, much less full sunlight. Because of the extreme ambient light range typical of commercial flight decks (0.1 to 8000 foot lamberts), it was not possible to use more than two levels of symbol brightness without a risk that the lower brightness symbols would disappear under some conditions. Extensive laboratory and simulator testing was done to develop symbols that were easily recognized and correctly associated with the information they represented. The primary coding tools available were symbol shape, line type, and text. Simple shapes were used as much as possible to minimize display clutter, improve readability in turbulence, and control the amount of drawing time required. Even with all this attention, the displays became quite busy.

In the late 1970s, color display technology improved enough to warrant their consideration for flight deck use. The general presumption was that color would simplify the information coding problem. In fact, coding was a secondary reason for using color, the primary objective was to separate the various classes of information to operationally declutter the display.

Two characteristics of human color vision played a key role in establishing this objective. Color is recognized over a much larger field of view that the small zone of sharp visual acuity where details of shape will be perceived. This permits a different, and potentially quicker, search strategy to be associated with color information.

Color CRT displays are affected by bright sunlight in two ways. The contrast of the symbol against its background is reduced in much the same manner as with monochromatic displays. More subtly, the color of the sunlight mixes with the color of the symbol, shifting the hue and saturation that the pilot perceives.

Accurate recognition of color is marked by significant individual differences. Testing conducted by Boeing showed that no more than six colors (seven under certain conditions) could be discriminated by the full range of pilots having "normal" color vision under all anticipated brightness conditions. This finding is not as operationally restrictive as one might first believe. In fact, it is conveniently appropriate. Psychological research consistently reports that human beings will have the least difficulty dealing with memory-based codes of not more than 7 (±2) dimensions.

A third consideration was stereotypical meanings associated with various colors by different cultures. Red is widely associated with warning or alert conditions. Amber (or yellow) is often recognized as indicating some form of caution or heightened awareness. Other colors have more diverse cultural meanings and are, therefore, more suited to general grouping than detailed operational information coding.

Using color this way does not reduce the need for shape and line style coding. However, it does permit higher information density to be used without incurring a pilot performance penalty. Dissimilar redundancy in the coding can improve pilot confidence in the display and help maintain good performance under marginal operating conditions.

Eye Fatigue

The use of CRT displays introduced new opportunities for eye fatigue. To minimize this potential, several characteristics of the displays were carefully controlled. Eye fatigue results when the muscles controlling the eye are subject to overuse.

The muscles that change the shape of the lens respond to the sharpness of the edges in the image falling on the retina. For conventional mechanical displays, edge sharpness is very high. The manner by which a CRT image is created produces a Gaussian-like distribution of light across each line in the display. If line widths, along with phosphor dot arrangement and spacing, are not carefully selected, the resulting soft edges can cause excessive refocusing and eventual eye fatigue.

Laboratory testing with a variety of pilots revealed that the optimum line widths for color CRT displays were significantly wider than for monochromatic CRTs and that the desired widths varied with the color of the line. This latter finding appears, at least in part, to be related to the fact that misconvergence can cause color fringing along those lines composed of two or more primary colors.

Eye fatigue can also result from fixating in one location for an extended time. Fortunately the distributed nature of information in a modern flight deck encourages the pilot to change his point-of-regard frequently. When the large format CRTs were first proposed for the 767, there was concern that the novelty of the display along with the large amount of information they contained would result in much greater dwell time on these instruments than was true of previous displays. The original performance criteria for the displays included graphic symbology that could be interpreted quickly by the pilot. Eye

track records confirmed that dwell times remained quite similar to those associated with conventional displays.

A third potential source of eye fatigue is the apparent motion in a display caused by flicker. Rapid motion is a powerful means of attracting visual attention. This is true for any visual scene real or created. The motion attention response is so automatic, that it is not under the conscious control of the pilot in most situations. The human visual system's sensitivity to flicker is not uniform throughout the visual field. For most people, it is greatest in the peripheral region between 45 and 60 degrees away from the eye point-of-regard. In this region, the critical flicker fusion frequency generally will not be less than 45 Hz nor more than 62 Hz. This is significantly higher than in the foveal region where critical flicker fusion frequencies below 30 Hz are common.

Unfortunately the zone of greatest flicker sensitivity overlaps the location of the other pilot's displays in most side-by-side two-pilot flight decks. Thus the required refresh frequency for flight displays is set by the flight deck geometry. For displays used on the 757 and 767, the nominal refresh rate is 80 Hz. This is allowed to drop, under high data presentation conditions, to as low as 65 Hz. Below that frequency, a message appears alerting the pilot to the data overload condition and allowing him to deselect unneeded information.

Though glare and reflections do not cause eye fatigue directly, they are likely to be reported as such if the pilot becomes aware of them on a continuing basis. The use of an anti-reflective coating on the external surface of the display along with careful matching of the index of refraction for the various layers of the display face plate greatly reduces the opportunity for perceived reflections. Finally, the flight deck geometry is established to ensure that sunlight on the pilot's white shirt will not reflect off the screen and into the pilots' eyes in his normal seated position.

Attention to all of these details has resulted in displays that pilots regard as highly readable and with which they achieve consistently high performance. Future technology changes will likely alter the specific requirements characteristic of current displays. Even some of the areas of concern might change. However, by understanding the factors that influence both perceptions and performance, it will be possible to ensure that the next display technology evolution is at least as successful as the transition to CRTs has been.

Time Shared Information

When the primary display devices were mechanical, there were few opportunities to time-share display space. True enough, VOR course deviation, ILS localizer deviation, and possibly inertial cross track deviation can be displayed on an electro-mechanical HSI. However, most other electro-mechanical displays have a fixed information content and a fixed format for that information. The change to CRT displays presented the opportunity to change the conventional one display one function relationship. With this opportunity came the necessity of understanding the circumstances under which time sharing would alter pilot performance. The potential to improve performance is there, and along with it, the potential to degrade performance.

Clearly a complete understanding of all the tasks that might be affected by time sharing of information is the appropriate starting point. The simplest cases involve tasks that can be isolated. It helps if these tasks are done relatively infrequently and under very clearly identifiable conditions. Slightly more sophisticated cases involve a change in priority or importance for a task, or tasks, which are necessarily serial in execution. The greatest challenge occurs when one or more tasks can occur in parallel with any number of other tasks and the relative priority of the tasks is known only by the pilot.

The first question asked by the designer should be, is it desirable to time share information for this task. If task execution is continuous, or nearly so, the answer is obviously no. For infrequently executed and logically isolated tasks, the answer is probably yes. The vast majority of tasks fall between these two extremes. In these cases the answer depends upon the composite impact of the total information display requirements on the pilot and the means available to effect the time sharing.

The map displays on all Boeing airplanes incorporate manually selected time sharing for supplemental navigation data. This includes depiction of navaids, intersections, and airports other than those currently in use or formally defined as part of the flight plan route. Manual selection is used since the specific circumstances favoring use depend on conditions known best by the pilot. All information that is mandatory for proper execution and monitoring of the defined flight plan is presented without specific action by the pilot. For example, the navaids currently being used for navigation updating are shown whether or not navaids manual data has been selected. The same is true of the departure and destination airports and any intersections that are identified as waypoints along the route of flight.

A variety of performance data is available when the pilot takes deliberate action that indicates such data would be useful. For example, the normal procedure for changing altitude is to select the new altitude on the mode select panel and then initiate a climb or descent, as appropriate. The two actions generate a prediction of how far ahead of the current position the aircraft will be when the new altitude is reached. This prediction is shown as a green arc on the map display. Once the new altitude has been captured, the prediction is no longer meaningful and it is automatically removed from the display.

A similar strategy is used to support the temporary engine exhaust gas temperature (EGT) limit that applies during engine start. A red radial is shown on the EGT gauge at the start limit value from the time the start is initiated by the pilot until the start cycle is completed. If the start operation occurs while the airplane is in flight, additional information is needed to ensure that enough air flow is available to complete the start. In this case, appropriate information about the airspeeds necessary for an unassisted start are displayed near the primary engine indicators when the engine is not running during flight. If the airplane is not at a speed sufficient for an unassisted start, the need for cross-bleed assistance is shown directly on the appropriate engine rpm indicator.

The time sharing illustrated by these examples would not have been possible without the flexibility of a general purpose display device like the CRT. The obvious benefit obtained from the engine and performance time sharing discussed above is the heightened awareness of the time shared data that occurs during the interval when that data is significant to the pilot. The corollary benefit may not be so obvious but is, nevertheless, one of the fundamental operational reasons for considering time sharing. This benefit can best be illustrated by noting that the most effective displays are those kept simple. Every extra display element takes time to interpret and introduces additional opportunities for misinterpretation and error. Further, the errors will not be confined to the extra data. As noted in the section discussing evaluation, the presence of nearby symbols, particularly dynamic symbols, can be a significant enabling factor for error.

A human characteristic that points toward the desirability of simple displays is the notion of selective attention, or "tunneling." In essence, under certain conditions many people have a tendency to fixate on selected data or tasks and ignore others. The circumstances that trigger this phenomenon are highly individual; but excessive workload, high stress, fatigue, or fear are often precursors. The task that is attended to may or may not be the most appropriate for the existing circumstances. Indeed, if tunneling continues for any significant time, it is likely that the data that would aid the pilot in recognizing the need for a priority change has, itself, been biased by the lack of

attention. The simpler the normal displays are, the more likely they are to avoid the tunneling phenomenon. If tunneling does occur and the displays are kept simple, there is a greater chance that the pilot will see only high priority information.

Another aspect of human perception that may play a part in the decision to time share is our human tendency to see what we expect to see. If data are continuously presented and are normal for an extended time, it is likely that the threshold at which a pilot will recognize an abnormality exists, will become less precise. Many tools are available to deal with this characteristic. Most depend on some form of alerting triggered by a parameter exceeding a limit value. Two examples illustrate ways of dealing with this phenomenon.

Exhaust gas temperature (EGT) is a basic engine health parameter on most jet engines. As such, EGT is required to be displayed in the flight deck. It has no other operational use. The actual value of EGT varies with engine power setting and altitude in a rather complex way. Thus, over a typical flight, the pilot can expect to see the EGT value vary from some low value to quite close to the limit value. Thus, proximity to the limit is not necessarily a concern, but exceeding the limit is. The reliability of modern jet engines suggests that, on average, a pilot would see an over limit condition not more than once every few years. That represents many hours of seeing normal values for every case of an abnormal value.

Simple limit values can usually be sensed precisely and reliably by the instrumentation system. That is the case for EGT. Several elements of the EGT presentation change color when the established EGT limit is exceeded. The color change affects the EGT pointer, the related EGT digital readout, and the box drawn around the digital readout. Since the majority of the Engine Indicating and Crew Alerting System (EICAS) display is white-on-black, this change to red-on-black is highly visible. With the color change, there is no doubt that the limit has been exceeded and which engine has the problem.

There are three engine types available for the 767 and the 757. A common type rating was planned between the two airplanes. None of the engines have exactly the same values for their limits, but they are displayed in exactly the same way. Therefore, the pilot doesn't have to memorize a new number when transitioning between airplane types. Instead, he uses the display exactly the same way on both airplanes. This is one of the versatile things that can be done with a CRT instrument.

The secondary engine instruments present a slightly different challenge. In this case, there are five or more parameters per engine. The values of some of these

parameters are only subtly linked to the pilot's operation of the engine. They may or may not have limits associated with them. Fundamentally, these parameters are used for long-term engine performance assessment, for backup if a primary indication fails, or for maintenance assessment if abnormal engine operation is encountered.

These secondary indications are grouped on the lower EICAS display. The design of this display is such that the data can be turned off without loss of limit indication. The computer monitors track those parameters that have limits and pop up the appropriate information on the display if a parameter goes out of limits.

Recommended usage of this feature for most engine-airframe combinations is to have the lower display active during engine start and then to blank the display for normal flight operations. Of course the pilot should activate the lower display any time he wishes to check any of the secondary data. The flexibility of use of this feature allows airlines and pilots to tailor operations to fit their particular operating style. At the same time, the availability of the feature recognizes that it is unreasonable to expect that all pilots will be properly attentive to displayed information regardless of the circumstances and the quantity of data actively displayed.

All the time sharing discussed to this point has involved changes to the data content of an existing display. In all cases, the basic conceptual framework for each display remains intact. The most general form of time sharing involves conceptual changes in the content of the display. In the extreme, this could mean that the display surface is used sequentially for totally independent tasks involving completely different information.

Successful implementation of this type of time sharing requires careful attention to the details of all related tasks and for the circumstances under which switching from one task to another will occur. Recognizing and supporting all the task linkages that can occur, particularly those associated with non-normal operation, is a key prerequisite for success.

Selecting the various modes of a time shared display will be most successful if the conceptual model used to implement the switching, matches the pilots' understanding of system usage. For complex systems, this is a difficult task since the level of system usage understanding will likely be different from pilot to pilot. Understanding will also be different for a single pilot as his skill with the system evolves from novice to expert. For example, a tree-structured selection concept is often preferred during initial training but shifts as experience is

gained to a preference for direct selection, particularly for frequently used features.

Accommodating this shift can be accomplished in many different ways involving design or training or some combination. Deciding what is best in a particular application is a complex task. No one answer is correct for all situations. A thorough understanding of the tasks, and their criticality in relation to other flight tasks, is the best basis for initiating the decision.

Command vs. Situation-Prediction Displays

Most flight deck displays support continuous control tasks, decision-making tasks, monitoring tasks, or some combination of these. At the task level, the supporting display information can:

1) show the current situation;

2) show what should be done to accomplish an established goal, or

3) show what will happen if the current action is maintained.

These three types of information can be categorized as: situation, command, and prediction respectively. Various combinations of these data can be used to optimize support for specific tasks.

Situation data is fundamental to many monitoring tasks and most, if not all, decision-making tasks. Command information is often associated with high precision control tasks. Prediction information can be used with all three task types though usually it is not used alone but in conjunction with situation information.

Situation information has the broadest applicability across tasks. It often entails more information transfer to the pilot than other means. The minimum situation data to support the lateral control task would involve:

- o current airplane location with respect to the desired location,
- o airplane heading (or track angle),
- o airplane speed,
- o airplane bank angle,

- o limits associated with any of these parameters, and

- o any other applicable constraints.

Understanding all of these data places the pilot in an excellent position to recognize subtle deviations from plan or in expected performance. It also gives him the widest possible range of task execution strategies. At the same time, it requires considerable skill to correlate all of this information correctly and to select the proper control strategy. Even for highly skilled pilots, there are practical limits on how fast this task can be completed correctly.

Command information simplifies the information processing load on the pilot by integrating the relevant information into a new piece of information indicating how much control should be applied. By presenting to the pilot the difference between the computed desired control input and the actual input, he can see immediately what should be done. This greatly reduces the information processing workload on the pilot and reduces his response time essentially to that associated with simple eye-hand coordination.

There are several costs associated with command information. The reduction in processing load on the pilot means that his awareness of the situation is also reduced. Similarly the choice of execution strategy is handled by the command generator rather than the pilot. Where performance demands are high, these costs may be considered acceptable or they may be reduced by procedurally involving the other pilot in some portion of the task.

Predictive information, like command information, combines data to reduce the processing workload. However, while the command information is based on a predefined control strategy, predictive information is based on the existing control strategy. Furthermore interpretation of the prediction requires enough understanding of the situation to determine the suitability of the current control input. This explains why most predictive information is presented in the context of a situation display.

The 767 map display contains several predictions. Those associated with lateral maneuvering clearly illustrate the differences between prediction and command information. In determining how to maneuver laterally, the pilot has a number of decisions to make. One involves how much turn rate is needed and another involves how quickly to roll out of the turn. A command display would indicate how much bank to use for a pre-established turn rate by showing a bank command to the pilot at the appropriate time. Then, at the appropriate time for roll out, an opposite bank command would indicate how quickly the pilot should reduce the bank angle to re-establish level flight.

The corresponding predictive information on the map (see figure 10.6) consists of a variable radius circular arc symbol whose radius varies with the current turn rate. In this case, the pilot can see that he has selected the proper bank angle when the arc is tangent to the desired path or when it passes through the desired point ahead of the aircraft. A fixed straight line from the airplane symbol to the top of the display shows the path the airplane would follow if the turn rate were zero. The rate of closure between this symbol and the desired path line or target way point and this fixed line provide the position and rate information the pilot needs to select and control his roll out to level flight again. These predictions are very simple but very powerful.

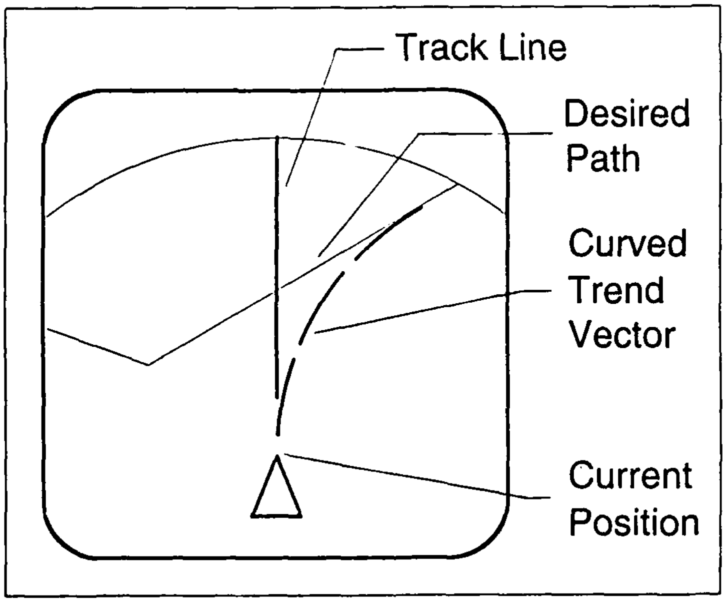


Figure 10.6 Variable radius circular arc symbol whose radius varies with the current turn rate. (original figure).

The length of the curved trend vector is proportional to the airplane ground speed. Gaps in the curved trend vector show where the airplane will be 30, 60, and 90 seconds ahead of current position. Of course the pilot can get some sense of speed from how fast the map information is moving beneath the airplane symbol. However, the fixed time intervals of the arc symbol provide the pilot with a relative time reference to use in interpreting the rest of the display information.

The predictive information does not directly tell the pilot when to maneuver nor does it demand a particular maneuvering strategy. The pilot must make these

decisions. In order to make them, he must have an understanding of the current flight situation. After a little practice with the predictive information, the pilot can make those decisions very accurately. Because of his interaction with the rest of the map information, good situation awareness is ensured.

Predictive displays are best suited to tasks where both deviations from some plan or standard and some form of rate information are involved. These displays are usually superior to other forms, where both control and a related monitoring task must be performed.

Which type of display is best? The answer is the one that most consistently and accurately enables the pilot to achieve the performance goals associated with the task he is doing. Here again is strong support for the necessity of understanding the task and the related information requirements before selecting the display format or symbology.

Future Display Issues

The broad acceptance of computer-generated data and the trend toward graphical user interfaces suggests that the flight decks of the future will contain more general purpose displays and that the pilots will expect to see much of the data presented in a graphic form. Technology trends indicate that flat panel displays may well replace CRTs as the display of choice for many applications. The detailed human factors issues associated with flat panel displays are quite different from those of the CRT since the image generation mechanisms are completely different. Though the technology details are different, the methodology for developing and evaluating such displays will remain consistent with the process outlined in figure 10.1. In the past, there has been a steady trend towards more and more data being made available to the pilot. Large format, computer-generated displays can readily overwhelm the pilot with information. Adherence to a structured process for evaluating pilot performance when using these displays will become increasingly necessary. Techniques such as time-sharing and adaptive selection of display information will be primary aids to the designer in coping with the information expansion. The certification issues raised by these techniques will need thoughtful consideration and debate.

Effective management of the rapidly expanding flight deck information system will require the cooperation of many people and organizations that support the pilot. A common understanding of both desired performance and actual performance along with the means to share this understanding across the industry will be very helpful. Human engineering plays a significant part in this process by providing a common understanding of the pilot and his performance to all of the participants in this endeavor.

Chapter 11

Workload Assessment

by Delmar M. Fadden, Chief Engineer-Flight Deck, Boeing Commercial Airplane Group

Workload assessment became a formal part of the certification of large commercial transports, with the adoption of Appendix D to FAR Part 25. While Appendix D identifies the need for such assessment it does not define the means. In retrospect, the fortuitous lack of rigidly defined methodology prompted considerable research and development that otherwise might not have occurred. The expansion of workload understanding and of the methods for assessing workload has enabled the industry to keep pace with the rapidly evolving character of crew workload over the last quarter century.

Operational differences in airplanes, such as the 737, 757/767, and 747-400, cause changes in the workload the pilot experiences. The nature of these changes has led to changes in the tools used to assess workload. On the 737, the workload of primary concern was the shift of system management

responsibility to the two pilots. The flying task assigned to the pilots did not change significantly between the 727 and the 737. The physical layout of the column and wheel, primary flight displays, and the cockpit windows remained very similar to the 727. The tasks that did change were those associated with engines and systems management. The engine management tasks were subtly different, reflecting the twin engine configuration of the 737. The systems underwent substantial change to bring them into conformity with the two-pilot operating concept.

By the time the 767 design was initiated, extensive experience had been obtained from a wide range of two-crew operations around the world. This experience confirmed the soundness of the basic principles underlying the design of systems for two-crew operation. However, airline desire for improved operating efficiency, coupled with the increasing complexity of the air traffic control environment, argued for significant enhancements to the primary flight information. A new flight management system concept was devised featuring cathode ray tube (CRT) flight instruments and digital computers handling many of the navigation, flight planning, and performance assessment calculations. These changes altered the pilots' tasks in ways that achieved improved efficiency and greater overall situational awareness. These changes produced corresponding changes in the pilots' experience of workload.

The 747-400 incorporates both the systems enhancements that had been pioneered on the 737 and the flight management capabilities first introduced on the 757 and 767. In addition, the primary instrument panel is modified permitting the use of larger CRT displays. Finally, a number of new information management features assist the pilot in coping with the increasing quantity of flight, engine, and systems information available. These changes, along with a complete redesign of the airplane systems, made it possible to change the crew size from three, as it had been on previous 747 models, to two. The workload concerns in this case focused on the integration effectiveness of the overall flight deck design.

This chapter reviews the evolving techniques that have been found useful for assessing workload in modern jet transports. Emphasis is placed on workload assessment in the early stages of design, since that is the time where quantitative workload data is the most effective in shaping the product. The techniques that have been developed to add structure to the subjective assessments of the evaluation pilots are described. Several issues that have significant effect on workload and the workload certification process are presented. The chapter concludes with a discussion of pilot error and a glimpse at future workload issues.

Workload Methodology

Commercial aviation, during the jet age, has established an excellent record for safety. The skills of many pilots have been a vital factor in that achievement. Nevertheless, when accidents do occur, history indicates that some type of pilot error will be involved in over 70% of the cases. Any work that leads to a reduction in the consequences of pilot error has the potential to improve the future accident record. While pilot workload, per se, has never been cited as the cause of an accident, there is a common perception that workload and error are related in some fashion.

Workload on a commercial airliner seldom, if ever, reaches the absolute limits of the flight crew. However, circumstances do arise which result in a significant elevation of workload. Whether or not such increases are large enough to cause concern about the potential for error is one of the reasons for doing workload assessment. The general relationship between workload and error is not well understood, even within the human engineering community. There is general agreement that error increases at both extremely low and extremely high workload levels. In between, evidence for any direct relationship is weak or nonexistent. Individual differences between pilots contribute to the difficulty of establishing a useful working relationship for workload and error. There appear to be significant variations in the level at which workload is considered extremely high or extremely low from one individual to another and, even, for the same individual under different personal and environmental circumstances.

Regulations applicable to commercial aircraft treat workload as a series of factors that must be considered for each of the primary flight functions. The workload factors, identified in Appendix D to FAR Part 25, constitute several of the key dimensions through which a pilot experiences workload. The characteristics describing these factors remain reasonably consistent for any one pilot across a variety of vehicles and flight conditions. Differences among individuals, however, tend to be large. The workload functions, also identified in Appendix D, encompass the major functional tasks normally assigned to the pilot. The details of these tasks, the related specific performance objectives, and the relative task priorities, vary considerably from one aircraft type to another.

Workload assessment plays a dual role in the design and development process. During the design cycle, workload assessment provides insights about the design that identify opportunities for improving the pilot interface. Workload assessment during the certification process provides a structured method for examining the various workload issues that are relevant to the particular aircraft type under scrutiny. Because it is very difficult to change the fundamental factors that establish crew workload after the airplane is built,

manufacturers place heavy emphasis on the selection and use of assessment methods that correlate well across these two roles.

The design development role argues for assessment methods that are both sensitive to detail and quantitative. The number, type, and timing of required tasks are important elements in determining how the design of the flight deck will influence the pilots' subjective experience of workload. Yet the pace of most development programs is such that workload assessment methods must be simple enough for timely application. Furthermore, since the entire airplane design does not approach maturity at a constant rate, the workload methodology must support assessments of isolated systems as well as assessments of the entire airplane.

For certification assessment, the diagnostic sensitivity of the workload method is less important than its overall vehicle applicability. The reality of certification in a social and political, as well as technical, environment means that particular attention must be paid to any unique or unusual features of the vehicle or its environment. Thus, the certification methodology must be flexible enough to adapt quickly to new tasks, new technologies, or new human performance concerns.

Since aviation progress is normally evolutionary, each new airplane type will contain a mixture of significant design changes and designs closely linked to previous airplanes. History during the jet age indicates that the elements of design undergoing the greatest change shift focus from one generation of airplanes to the next. It is, therefore, not surprising that the analytical methods that have been developed, depend on comparisons between the new design and existing designs having an established safety and operational performance record.

The multidimensional nature of the workload experience makes it unlikely that a single absolute workload scale will ever be developed. Indeed there is reason to suspect that creation of such a scale would be of little practical utility in the development of commercial cockpits. Instead, all current workload assessment techniques involve multiple measures, most of which depend on some form of comparison. The comparison will determine if the new design has the higher workload. Whether the difference is significant depends on the magnitude of the difference, the length of time the difference remains, and the phase of flight when the difference occurs.

Commercial Aircraft Workload

Commercial aircraft workload can be divided into two broad regimes: normal and nonnormal. The former constitutes all the tasks associated with planned operation of the aircraft, including:

- o all allowable flight operations,
- o all certified weather operations,
- o certified minimum crew size,
- o selected equipment unavailability under the minimum equipment list, and
- o normal flight operations following probable equipment fault or failure conditions (exclude tasks associated directly with management of the fault or failure).

Normal workload presumes compliance with all operating and performance requirements along with adherence to all restrictions, limitations and established policies. Under nonnormal conditions, strict compliance with normal operating requirements can be relaxed, as long as aircraft or personnel safety is not further compromised. In addition, through appropriate coordination, it may be possible to relax adherence to certain externally imposed restrictions or performance standards. Such relaxation plays a significant part in mitigating additional workload that might otherwise accrue from nonnormal events.

All remaining tasks are considered nonnormal. Both the consequences of occurrence and the probability of occurrence are considered in determining which nonnormal tasks are identified with specific procedures in the operational documentation and the training the pilot receives. During design, assessments are made of all possible ways in which safety hazards can occur. In this manner, the relevance of every nonnormal event is determined. Experience shows that particular attention is needed for events that are associated with:

- o other than normal flight conditions,
- o incapacitation of a required crew member,
- o management of equipment fault or failure conditions,

- flight operations subsequent to improbable equipment fault or failure conditions, or

- flight operations following combinations of faults and nonnormal events.

An important aspect of nonnormal workload management concerns the design of equipment and procedures that minimize the consequences of failures on subsequent aircraft operations. This focus has the obvious benefit of reducing the aggregate workload, but, what is more important, it also reduces the opportunities for error that would accompany a sustained change in procedures. This principle is embedded in the systems design for Boeing airplanes and has produced many nonnormal procedures that are independent, time-limited task sequences. This results in getting the pilot back to normal flight operations and normal procedures very quickly for most first failure conditions.

While care must be exercised to avoid unnecessary workload buildup, staying well below the pilot's maximum workload capability is a relatively straight-forward task to accomplish on a commercial flight deck. Important as it is, attention to task loading alone is not sufficient to ensure an error-tolerant flight deck. The timing of tasks plays a significant role in determining what opportunities for error may be encountered. Thus, it is recognized as desirable to organize the normal task loading with the following timing-related guidelines in mind:

- it should be possible to interrupt any procedural task sequence at any point to accomplish time or event-driven actions,

- abrupt changes in normal task loading should be avoided, particularly, during the departure and arrival phases of flight,

- the need for precisely timed tasks should be minimized,

- where task start time constraints are necessary, task completion time requirements should be relaxed,

- similarly, where task completion time constraints exist, the start time requirements should be flexible.

Rigid application of these guidelines is not necessary, but deviations should be treated as circumstances meriting special attention.

Workload Assessment Scheduling

Figure 11.1 shows a typical workload assessment program. Workload assessment is initiated early in the process so that the results can be used in optimizing the design. A typical airplane development program at Boeing usually takes five to

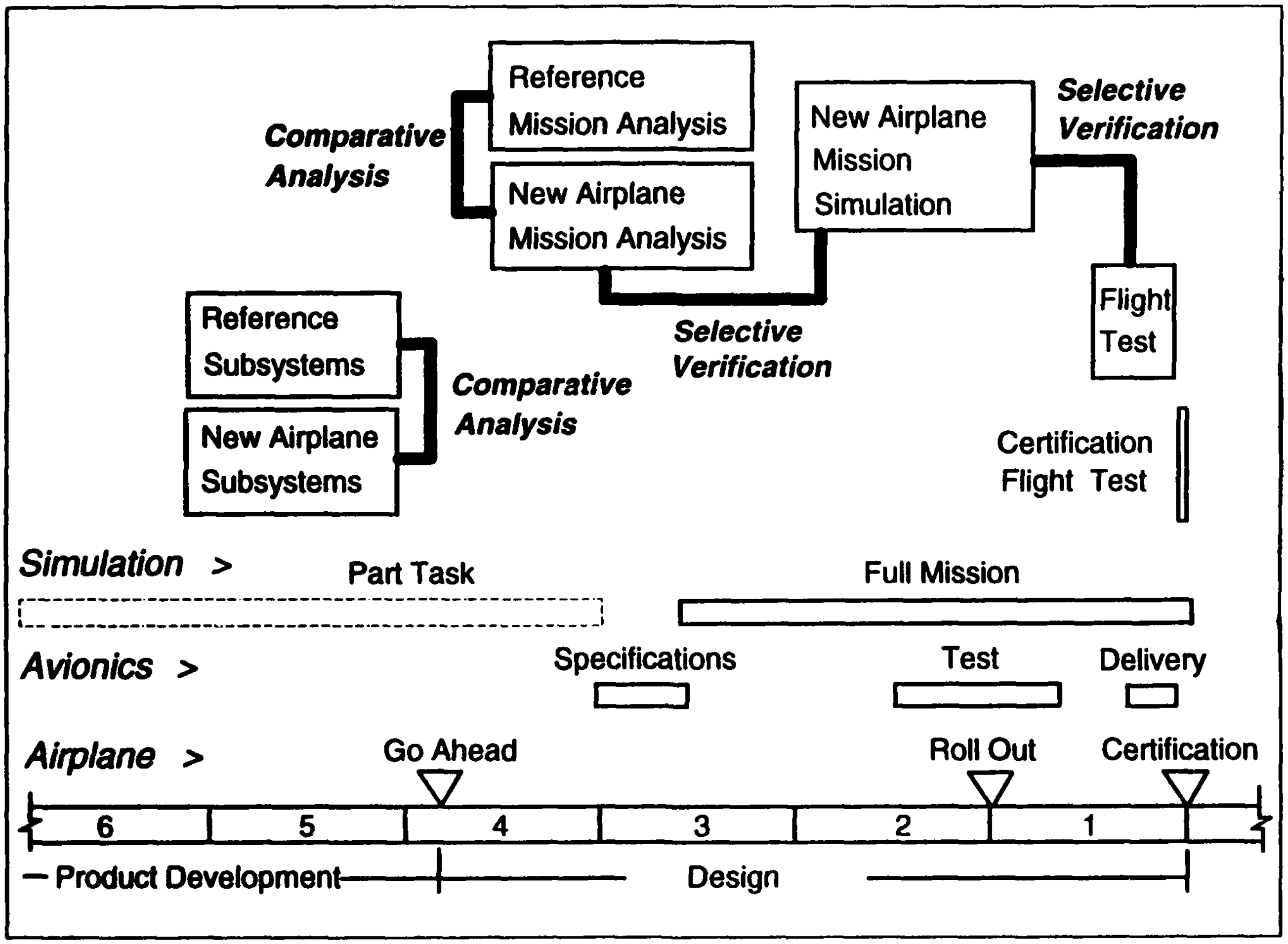


Figure 11.1 A Typical Five-Year Workload Assessment Program.

six years. The fundamental decisions that shape the basic airplane itself are frequently made in the first 12 to 24 months of the design activity. Structured workload assessments usually begin about 50 months before certification. The assessment tools selected at this point provide useful insight even though many of the details of the design are not yet finished. Where a reasonable degree of task similarity exists, comparative analyses based on these tasks can provide an

anchored reference. Where the task or the information presented is new and cannot be quantitatively linked to previous designs, some form of laboratory assessment, part-task simulation, or even experimental flight test may be necessary; particularly if the task is important to the safety or operating success of the airplane.

The costs of using these tools, particularly simulation or flight test, are not limited to dollars but extend to the time and human resources they absorb. Since committing resources reduces their availability for other developmental work, those issues selected for this type of testing are carefully considered and prioritized.

The first step in many new airplane workload assessment programs is a comparative analysis of the internal airplane systems; electrical power, hydraulic, pneumatic, environmental control and fuel being the most important. This analysis depends on system knowledge but requires little detail about events external to the airplane. Any impacts associated with external events or inter-system effects will be incorporated in subsequent analyses. Negative or neutral analytic results indicate where to focus further design attention. As with all analytic methods, these analyses provide visibility based on known or hypothesized relationships. Additional testing must be done when the possibility of unanticipated relationships between design elements or crew tasks cannot otherwise be reduced to an acceptable level.

During design of the 767, the analytic workload assessment process resulted in two additional design optimization cycles for the hydraulic system and one added cycle for the pneumatic system. These cycles occurred well before hardware was built at a time when significant design flexibility remained. Similarly, the fuel system of the 757 was changed from a five-tank to a three-tank configuration based on workload considerations. The fuel tank issue is particularly interesting because it illustrates the complexity of achieving truly effective designs.

Fuel is a major element of weight in a long range airplane. The distribution of that weight in the wing affects the stress each portion of the wing will experience during flight. The structural weight of the wing is directly related to these stresses. Naturally, the more the structure weighs, the more fuel must be carried to lift the extra weight. It is advantageous to reduce the bending stress by having more weight remain within the outboard portion of the wing than the inboard portion as fuel is burned during flight. Consideration of these factors for an airplane the size of the 757 suggested that the best structural design solution would involve five fuel tanks: two outboard, two mid-wing, and

a center tank. However, such a system would make necessary additional routine actions to manage the flow of fuel to each engine.

On a three-tank system, the center tank pumps normally operate at higher output pressure than the wing tank pumps. This ensures that center fuel will be used first. Operating procedures can be kept very simple; turn on all pumps before take-off and turn off the center pumps when center fuel is exhausted. Managing a five-tank system is more complicated for the pilot, unless a system is added to sequence the fuel automatically. Considering the criticality of the fuel system and the additional complexity that would be necessary to compensate for new failure modes, such added automation would result in an increase in electronics weight, several new nonnormal procedures, and increased maintenance requirements. Several design iterations addressed each of these issues and resulted in a revised fuel system that achieved lower total weight and the simplicity of a three-tank design. Reaching this decision required agreement across several, otherwise independent, functional groups within the design organization, the regulatory organization, and the airlines, thereby adding considerable time and effort to the design. The in-service results suggest that the effort was worthwhile.

This example points out how important the early workload estimates are. Redesigning the tank layout would not have been practical had the workload assessment been delayed until a full mission simulation or a flight test vehicle was available. In this case, the workload concern was identified by the manufacturer who then took timely steps to resolve it. Had the issue been a regulatory concern, it would have been equally important to identify early.

Workload Assessment Criteria

Should analytic workload techniques be used for certification? It is convenient for the manufacturer if they are, because the manufacturer has already applied them, and based a design upon them. If the regulatory agency and the manufacturer both agree on the scope and validity of such methods, then they can be highly useful.

Boeing starts with a subsystem analysis program called Subsystems Workload Assessment Tool (SWAT). The SWAT program assesses both normal and non-normal procedures. The primary purpose of this program is to relate the operating procedures, the display and control devices, and the geometry of the cockpit using a common measure. The subsystem analyses are not related to a specific mission so all the normal and nonnormal procedures are accomplished serially. The analysis encompasses time and motion assessments for hand and eye tasks. Such ergonomic data is essential for ensuring that displays and

controls are properly located within the system panel. Time and complexity assessments for aural, verbal, eye, hand, and cognitive tasks are also examined. The complexity score is a method of estimating the mental effort related to gathering information. It characterizes the information content of the displays and the number of discrete operating choices available to the pilot using a logarithmic measure (BITs). SWAT generates summary statistics for each system and for all systems.

Table 11.1 is a systems workload data summary comparing 767-200 normal inflight procedures with those for the 737. The data reflect that the 767-200 systems require that the pilot only switch off the two center tank fuel pumps when the center tank is depleted. The 737 requires a few more hand and eye tasks.

Table 11.1
Subsystems* Workload Data Summary
Normal Inflight Procedures (Boeing, 1982)

	Eye Activity Channel			
	Motion (Deg.)	Time (Sec.)	BITs**	Tasks
737	212	32	32	14
767-200	1	3	2	2
	Hand Activity Channel			
	Motion (Inches)	Time (Sec.)	BITs**	Tasks
737	17	17	8	8
767-200	1	3	2	2

* Combines results for electrical, hydraulic, ECS, and fuel subsystems.

** The BIT score derives from the classic definition of information.

Table 11.2 is a similar systems workload data summary comparing all nonnormal inflight procedures for the same aircraft. This table provides a gross check on the overall effect of nonnormal procedures. If any of the 767-200 statistics had exceeded the comparable data for the 737, that would have been an indicator that additional investigation is essential. To understand how individual systems fare in the comparison, it is necessary to examine individual systems data at a more detailed level. Interpretation of these data requires thorough knowledge of the system operation and the intended pilot interface.

Figures 11.2 and 11.3 summarize workload evaluation results for various airplanes under normal and nonnormal procedures, respectively. The two-crew 747-400 evaluation shown in the top two graphs of figure 11.2, for example,

Table 11.2
Subsystems* Workload Data Summary
Non-Normal Inflight Procedures (Boeing, 1982)

	Motion (Deg.)	Eye Activity Channel Time (Sec.)	BITs**	Tasks
737	2214	330	348	169
767-200	1348	183	297	126

	Motion (Inches)	Hand Activity Channel Time (Sec.)	BITs**	Tasks
737	458	183	140	88
767-200	355	154	134	81

* Combines results for electrical, hydraulic, ECS, and fuel subsystems.

** The BIT score derives from the classic definition of information.

has a lower average number of tasks and lower average time to complete the tasks than the three-crew 747-200. These graphs show that the 747-200 results are similar to the results for the 737 and 767. These comparisons give the designer an initial indication of how the workload associated with a new design will compare with other airplanes. The *normal procedure eye analysis* (upper left graph) shows that the 747-400 has a larger average number of tasks, but that on average, the tasks take less time to do than on the 737. The goal in this case is ensuring that total task time is similar to the total task time of another airplane having a good operating history. The 737 has been used as a reference by Boeing since the development of the 767, because the 737 has an excellent safety record and is flown by more customers, in more environments, using a wider diversity of pilots, than any other Boeing airplane. Experience indicates that the 737 is highly tolerant of pilot error and that it supports many different operating strategies.

A workload assessment summary for nonnormal procedures is shown in figure 11.3. Numeric totals are not particularly interesting by themselves because, even under the worst of circumstances, the pilot will use only a small percentage of the nonnormal procedures at any one time. Minimizing the number of tasks per procedure is considered desirable. While the 767 nonnormal workload is consistently the lowest, the corresponding 747-400 workload is significantly closer to that of the 737 on these logarithmic graphs than to the three-crew 747-200.

In successfully reducing workload, designers can establish circumstances where the crew has limited opportunities to experience certain events. If the crew

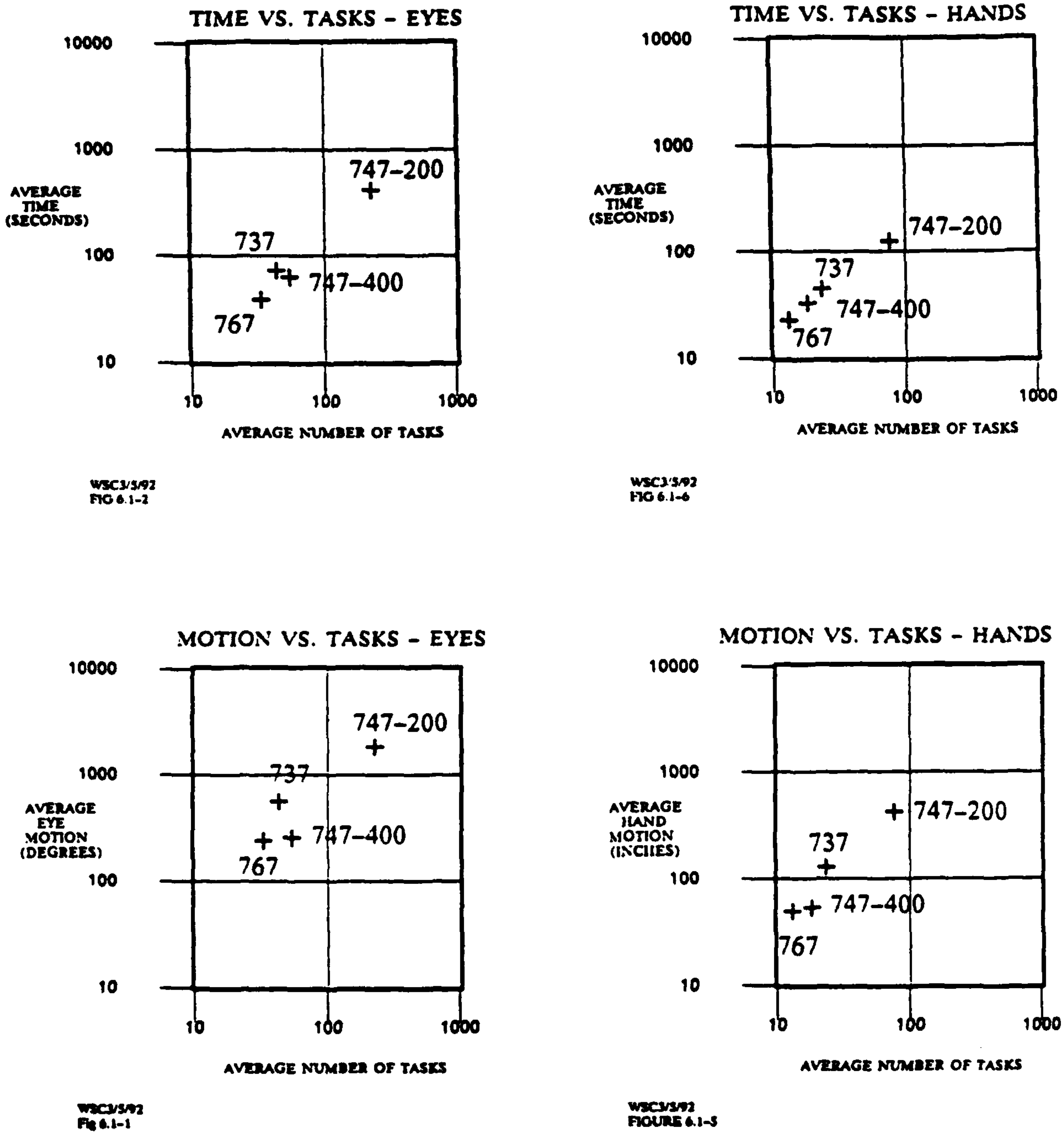


Figure 11.2. Systems Normal Procedures Workload Results for Various Airplanes. (Boeing, 1989)

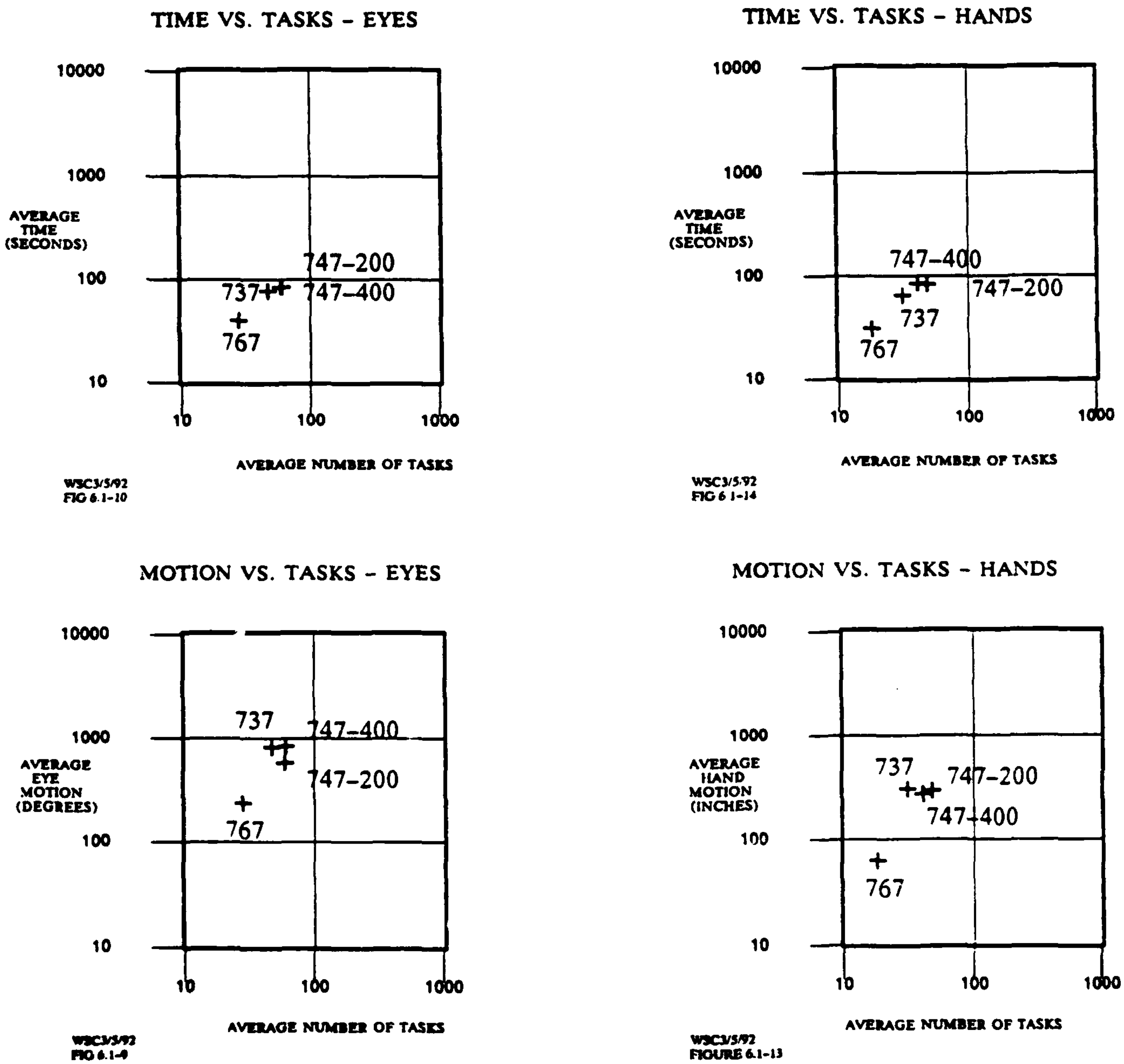


Figure 11.3. Systems Nonnormal Procedures Workload Results for Various Airplanes. (Boeing, 1989)

response to such an event requires proficiency at a physical skill, a mental data manipulation, or a complex decision, then some alternate means for developing and maintaining that proficiency may be required. It is certainly a poor trade-off to sacrifice achievable reliability and efficient operations simply to retain skills. As an example of the trade-off, the CRT displays on many newer airplanes are two to four times more reliable than the horizontal situation indicators (HSIs) on previous airplanes. As a result, the pilot doesn't have to use the standby instruments very often. The CRT presents data in a map format that is much easier to interpret than the symbolic presentations of the standby navigation instruments. Most pilots would say this is a good trade-off. However, this trade-off means that when the pilot must use the standby instruments, it is likely he will be less proficient with them than would have been the case on previous airplanes. Of course, simulator training can, at least partially, compensate for the loss of line exposure to the actual condition. If these possibilities are recognized early in the design process there may be other options. The designer may be able to design standby instrument flight procedures that are tolerant of a reduced skill level or allow for a longer transition period during which the pilot regains the needed skill.

Timeline Analysis

Once all the individual systems and panels are defined by hardware, functional description, operating procedures, and layout, the assessment process can be expanded to incorporate realistic operational scenarios. This permits quantitative evaluations of issues related to panel location within the cockpit, multiple system operations, and, most importantly, the time criticality of functions.

Time is one of the key dimensions of workload. Is sufficient time available for the pilot to complete all the tasks necessary to operate the airplane efficiently and safely? Timeline analysis is a structured methodology for examining this question. The fundamental equation for timeline analysis is the ratio of the time it takes to complete a task to the time available for the task. This sounds like a very simple idea. In practice there are many issues that must be addressed to accomplish the analysis. At the point in the development process where timeline analysis is first done, actual operating hardware is not yet available. Estimates of the time required to complete each task must be made. When hardware or a suitable simulator is available, the time estimates can be checked and appropriate adjustments made to the analytic data.

Timeline analysis is accomplished by examining what the pilot does in every 300-second (5-minute) time block along the course of an entire flight. The average workload over the entire flight is of little interest because the workload during departure and arrival is much higher than that during the, often much

longer, cruise phase of the flight. Consequently, statistics are focused on the arrival and departure phases of flight or on each of the 300-second blocks. Four separate channels of activity are examined: visual (eyes), motor (hands), aural, and verbal. Modal initiation and execution times for each task are recorded and each task is assumed to require 100% channel capacity for the duration of that task. Tasks are shifted as necessary to avoid overlap. Recent research results indicate that the 100% channel capacity assumption is significantly more conservative than necessary. However, it has proven useful, in the relatively benign workload environment of commercial aviation, by ensuring early identification of any brief periods when the assumption might be violated. It also avoids the necessity of collecting data justifying the selection of a lesser percentage.

The decision to keep the four channels separate has a similar expediency basis. From the design point of view, knowledge of the specific channel workload is essential if any adjustments are required. Thus, a combined statistic would serve only as an intermediate step toward getting to the specific channel workload. Combining the channel workload data immediately raises the question of the basis for the combination. With the exception of the aural-verbal pair, experience indicates that all the pairings can overlap successfully most of the time. The circumstances where complete overlap may not work appear to involve task events unfamiliar to the pilot or tasks of unusual complexity. The idiosyncratic nature of these circumstances makes a rule for identifying them difficult to develop and even harder to defend. The reason usually given for wanting a single workload number is to simplify the decision of whether the overall workload is acceptable. The lack of a firm basis for combining the channels has led Boeing to focus on the individual channel statistics.

Timeline analysis provides visibility of both dwell time and transition time. Dwell time is the time taken to read or operate the specific control or display device; for example: adjusting a control, reading information from a display, entering a way point name, or selecting a new switch position. Transition time is the time taken to switch from one activity to another. Examples of transition time are: moving the eye-point-of-regard from one display to another, moving the hand from the control column to the throttles, or changing from looking outside the cockpit to focusing on the instrument panel.

Comparing dwell time and transition time data for different flight decks provides useful information about the effectiveness of a particular design. If the dwell times for the design are high, then the system designer needs to consider using alternative display formats or control devices. If the transition times are high, the flight deck designer is prompted to examine alternate physical arrangements of the various controls and displays. Table 11.3 is a flight

procedure workload data summary for a Chicago to St. Louis flight depicting dwell and transition times for eye and hand activities. These are total dwell and transition times needed for the entire flight. These particular data were generated early in the 767 development as a gross indication of the design progress.

Table 11.3
Flight Procedure Workload Data Summary
Chicago to St. Louis Flight Totals (Boeing, 1979)

	Dwell Time (Sec.)	Transition Time (Sec.)	BITs*
Captain			
Eye Activity Channel			
737	550	24	2,271
767-200	372	20	1,811
Hand Activity Channel			
737	199	119	629
767-200	161	91	370
First Officer			
Eye Activity Channel			
737	510	30	2,423
767-200	331	26	1,844
Hand Activity Channel			
737	274	181	895
767-200	196	136	488

Other useful statistics generated by the timeline analysis program include the average amount of dwell time spent on a particular instrument and the probability of transitioning between various instrument pairs. Samples of these statistics are shown in Table 11.4. These two statistics are very useful in developing the most effective flight deck layout. Where these statistics depart significantly from those associated with current airplanes, the designer has reason to conduct more detailed studies.

The next two tables show the activity demands on the captain and first officer during each five-minute block of the one-hour, Chicago to St. Louis, flight. The total flight time is divided into 300-second (5-minute) blocks beginning at brake release and the time demands during each interval are shown as a percentage. The purpose of this form of data presentation is to examine the distribution of workload throughout the flight. Several characteristics are of interest in these tables. While none of the following trigger levels should be

considered a limit, exceeding of any of these levels is sufficient reason to conduct a detailed analysis of the activities within the interval. The results of the analysis will indicate whether the activities in the interval warrant adjustment.

Table 11.4
Flight Instrument Visual Scan
Dwell Time and Transition Probability Summary (Boeing, 1979)

	Average Dwell Time (Seconds)	
Instrument	Takeoff/Climb	Descent/Land
ADI	1.17	1.11
HSI	0.81	1.05
Airspeed	0.64	0.68
Altimeter	0.47	0.50
	Average Transition Probability	
Instrument Links	Takeoff/Climb	Descent/Land
Airspeed to ADI	0.90	0.86
Altimeter to ADI	0.87	0.79
HSI to ADI	0.78	0.80
ADI to Airspeed	0.25	0.23
ADI to Altimeter	0.36	0.28
ADI to HSI	0.31	0.36

Representative time-demand workload trigger levels are:

- o interval workload greater than 25%,
- o workload increase greater than 10% of total for consecutive intervals,
- o workload greater than the reference airplane for two consecutive intervals,
- o interval workload greater than 5% of total above the reference airplane.

Table 11.5 shows the visual activity time demands, while table 11.6 shows the corresponding data for motor activity time demands on the same flight. The flight scenario for this mission begins with a takeoff from Chicago (O'Hare) and a planned instrument departure. Once airborne, ATC provides radar vectors until the airplane is above FL240 when responsibility for "normal navigation" is returned to the pilot. Now the pilot returns to the cleared flight plan proceeding toward St. Louis. The cruise segment of the flight lasts for about five minutes after which the crew begins a standard arrival into the St. Louis area. The cleared arrival routing is different from the original flight plan.

Table 11.5
Line Operation Visual Activity Time Demand (Boeing, 1982)

Average Percent of Time Available Devoted to Visual Tasks*

Phase	Time Interval in Seconds	Captain		First Officer	
		767	737	767	737
Takeoff Climb	1 - 300	23	28	19	30
	301 - 600	9	19	8	20
	601 - 900	3	4	3	5
	901 - 1200	4	8	3	12
Cruise	1201 - 1500	10	16	5	9
Descent Land	1501 - 1800	12	24	10	13
	1801 - 2100	15	12	12	19
	2101 - 2400	9	13	17	13
	2401 - 2700	18	26	10	13
	2701 - 3000	8	15	5	12
	3001 - 3300	15	17	13	15
	3301 - 3600	5	10	14	17

*Excludes flight path control and outside watch.

During the descent, there is a runway change at St. Louis (Lambert). A thunderstorm on the descent flight path requires a detour. Finally, the visibility at St. Louis is low enough to require a precision instrument approach. Both airplanes are flown using the equipment provided on their respective flight decks. The performance of each airplane dictates the exact timing for the various events that occur. The data for the 767 generally indicates lower time demands than for the 737. The motor demands, in particular, are lower except early in the descent phase where the 767 pilots are receiving and programming the new arrival routing on the FMC-CDU. The 737 pilots have to respond to the revised routing as well, but without an FMC, they must wait to set their equipment until the airplane reaches the various maneuvering points in the procedure. This points out one of the advantages of having a flight management system: the ability to move selected tasks away from the later, lower altitude, portions of the flight path. As has been pointed out by many people, the introduction of new flight deck systems does not necessarily result

Table 11.6
Average Time Devoted to Motor Tasks* (Boeing, 1982)
(Percent of Available Time)

	Time Interval in Seconds	Captain		First Officer	
		767	737	767	737
Takeoff Climb	1 - 300	17	20	15	27
	301 - 600	6	11	10	21
	601 - 900	3	3	7	8
	901 - 1200	4	4	3	8
Cruise	1201 - 1500	8	9	6	11
Descent Land	1501 - 1800	9	11	14	11
	1801 - 2100	10	5	7	13
	2101 - 2400	4	6	20	12
	2401 - 2700	9	11	11	12
	2701 - 3000	3	5	4	12
	3001 - 3300	4	11	11	15
	3301 - 3600	7	10	<1	3

*Excludes flight path control and outside watch.

in lower total workload. Often the objective of a new system is to shift workload from one phase of flight to another. This is particularly true where routine involvement of the pilot is necessary to maintain the proper level of situational awareness. The flight management system is just such a system. By storing and displaying the flight plan before it is needed, the pilot is given the option of performing some tasks at a time of his choosing rather than having the task timing be established by the position of the airplane. While unexpected external events may occasionally reduce the value of this option, the data in tables 11.5 and 11.6 show the option can have an overall positive effect on terminal area workload.

The same interval-based time demand workload data can be shown graphically making the comparison somewhat easier. Visual channel timeline analysis data for the 747-400 is shown graphically in figure 11.4. Here again, the reference airplane was the 737 and the mission was a flight from Chicago to St. Louis.

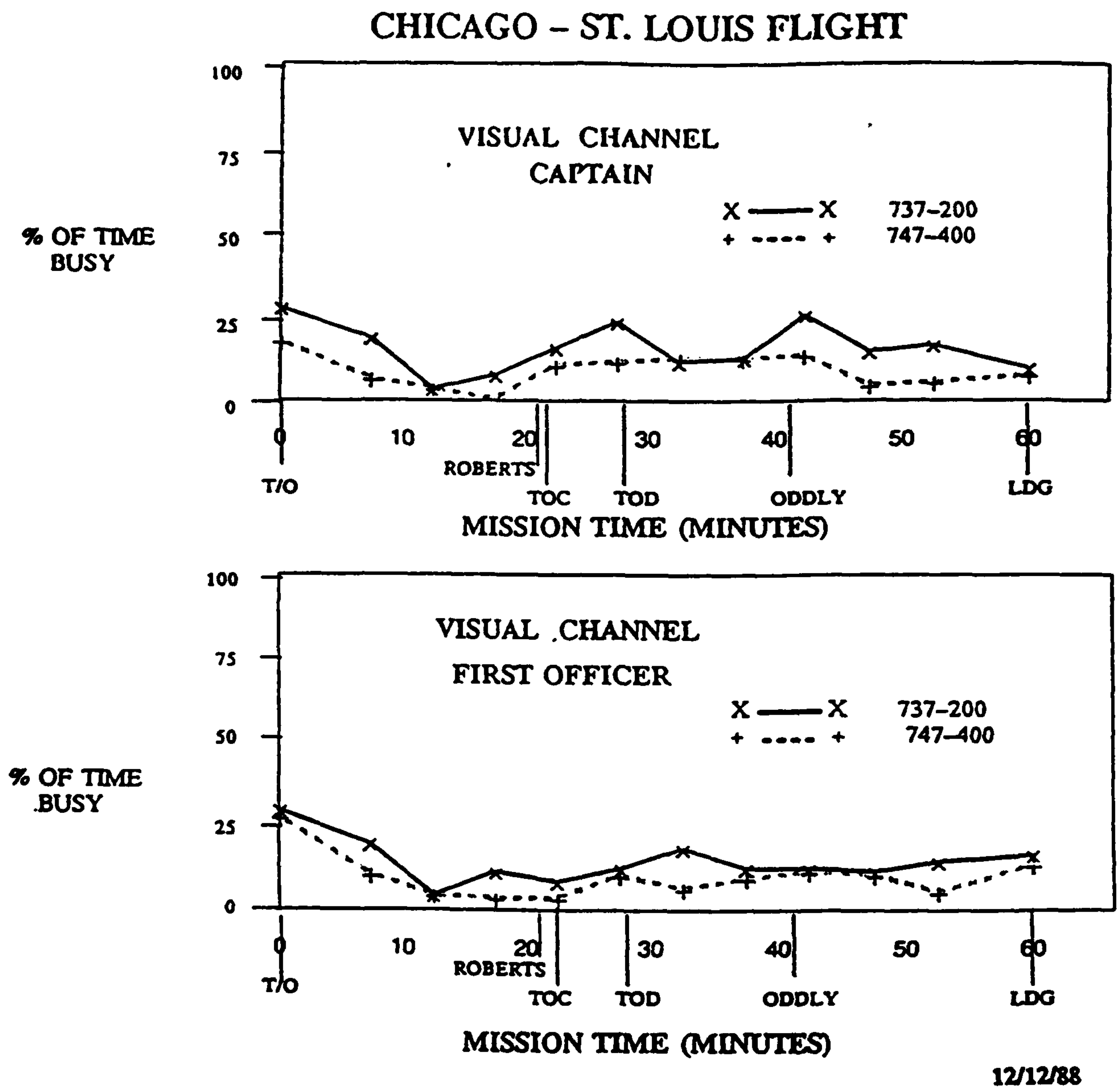
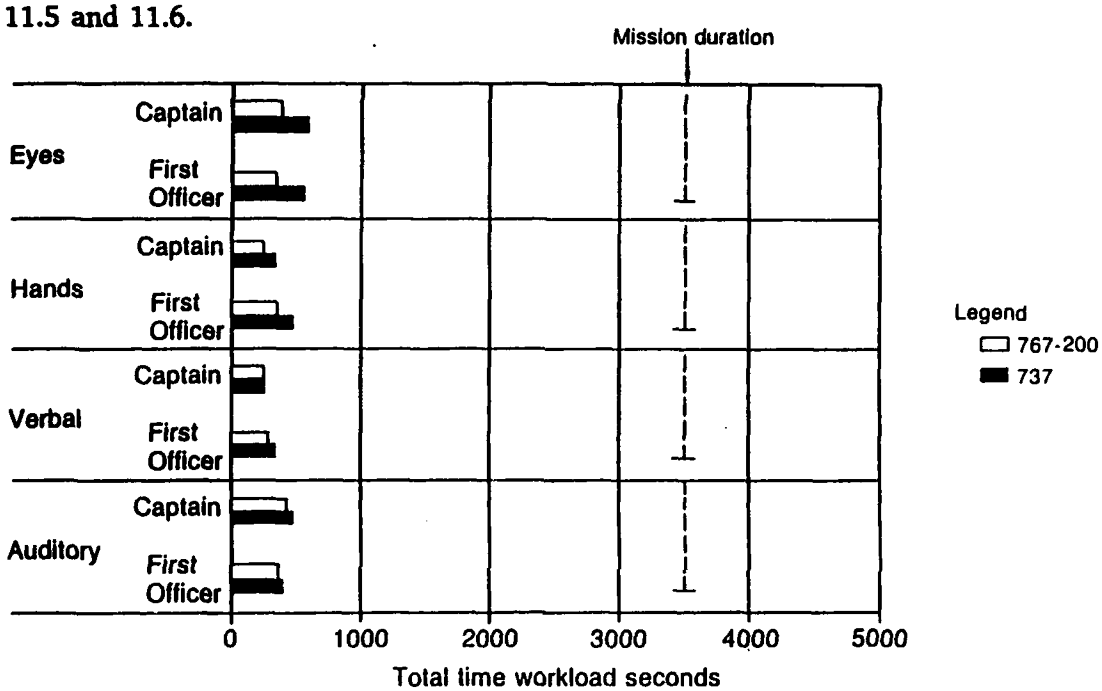


Figure 11.4 Mission Profile Visual Activity Time Demand, 747-400 and 737-200. (Boeing, 1988)

The final type of timeline analysis is a comparative plot of the total workload time for each of the four channels: eyes, hands, verbal, and auditory. An example is shown in figure 11.5 for both the 767-200 and the 737. The white bar represents the 767-200. The black bar represents the 737. This figure involves data for the same flight scenario used to generate the data in tables 11.5 and 11.6.



Analysis based on takeoff brake release at Chicago (O'Hare) to touchdown at St. Louis (Lambert)

Figure 11.5 Mission Activity Channel Time Demand Summary. (from Boeing, 1981)

Task-Time Probability

Another technique for examining task demands on the crew is called task-time probability. This method estimates the probability that the pilot will be busy with a task at each point along the flight. Since the method is probabilistic, it is possible to account for a range of pilot performance. Each task is associated with separate initiation and execution times, as was true for timeline analysis. However in this case, instead of being assigned discrete values, these two times are assigned probability densities. Tasks are allowed to overlap. Task-time probability is computed for each one-second interval along the flight path. The probability density functions are centered on the modal initiation and execution times. As a first estimate, a nonsymmetrical, triangular distribution is assumed unless more specific test data are available to support a different distribution. The nonsymmetrical distribution for task initiation or completion

Table 11.7
Probability of Being Busy with a Visual Task* (Boeing, 1982)
(Root-Mean-Square Probability)

	Time Interval in Seconds	Captain		First Officer	
		767	737	767	737
Takeoff Climb	1 - 300	.44	.49	.38	.52
	301 - 600	.26	.40	.25	.43
	601 - 900	.15	.19	.12	.22
	901 - 1200	.16	.26	.12	.30
Cruise	1201 - 1500	.29	.37	.19	.27
Descent Land	1501 - 1800	.31	.47	.30	.34
	1801 - 2100	.35	.32	.32	.42
	2101 - 2400	.27	.33	.37	.31
	2401 - 2700	.38	.47	.26	.31
	2701 - 3000	.25	.35	.16	.31
	3001 - 3300	.35	.38	.32	.35
	3301 - 3600	.20	.30	.36	.40

*Excludes flight path control and outside watch.

times recognizes that many flight tasks have either constrained starting or constrained ending times. The nonsymmetrical execution time distribution accounts for small variations in individual performance for highly skilled behavior and larger variation in performance where behavior is less skilled or involves more conscious effort. Examination of keyboarding test results using a number of different military pilots indicates that, at least for some tasks, the two distributions may not be entirely independent. Results show that the pilot who is slow executing a task is also likely to be slow initiating the task.

The value of this method is not how accurately the density functions characterize the pilot population but rather the insight that can be gained into interactive system performance at a point well before test hardware is available. The task-time probability statistics can be combined into the same five-minute blocks that were used for the timeline analysis. The various activity channels remain separate for the same reasons as were discussed in the timeline analysis section. For each channel, the second-by-second probabilities are combined into

a single number representing the root-mean-square probability statistic for each five-minute interval. Table 11.7 shows the root-mean-square probability of being busy with a visual task during each of the five-minute blocks of the same Chicago to St. Louis flight that was characterized in Table 11.5. Similarly, Table 11.8 depicts the root-mean-square probability of being busy with a motor task.

Table 11.8
Probability of Being Busy with a Motor Task* (Boeing, 1982)
Root-Mean-Square Probability

	Time Interval in Seconds	Captain		First Officer	
		767	737	767	737
Takeoff Climb	1 - 300	.39	.42	.35	.50
	301 - 600	.21	.30	.29	.44
	601 - 900	.15	.18	.22	.27
	901 - 1200	.15	.16	.14	.22
Cruise	1201 - 1500	.26	.26	.23	.31
Descent Land	1501 - 1800	.28	.31	.37	.32
	1801 - 2100	.28	.21	.23	.34
	2101 - 2400	.19	.22	.41	.30
	2401 - 2700	.27	.29	.30	.28
	2701 - 3000	.16	.17	.19	.31
	3001 - 3300	.18	.29	.31	.35
	3301 - 3600	.26	.31	.00	.16

*Excludes flight path control and outside watch.

Workload assessment using timeline analysis and task-time probability analysis is usually accomplished before there is a mission simulator in operation. As soon as a simulator is operating, the key task is to examine those segments of the flight where the analysis suggests that workload will be the highest. Simulator results can then be used to update the analysis. Spot checks in the low and medium workload segments provide increased confidence in the analysis and provide the opportunity to uncover any performance characteristics that were unanticipated.

Unless unresolved questions remain after the simulator testing, it should not be necessary to conduct instrumented flight tests simply to verify the analysis. Flight testing for quantitative time demand workload is extremely difficult to accomplish and is easily confounded by external circumstances beyond the control of the test conductor.

Pilot Subjective Evaluation

Quantitative workload testing in the actual airplane is much more difficult than in the simulator. The single biggest contributor to the difficulty is the unpredictability of the actual flight environment. At the same time, the actual flight environment improves the pilot's conscious sensitivity to variations in his experience of workload. That sensitivity can be focused and standardized using a well designed, structured workload questionnaire. Sample pages from a Boeing questionnaire for assessing pilot workload on the 757/767 airplanes are shown in Figures 11.6 to 11.10. By completing the questionnaire, the evaluation pilots indicate their experience of workload while operating either airplane. The specific workload functions and factors are related to those identified in FAR 25, Appendix D.

The questionnaire is structured to ensure that the pilot specifically thinks about the departure and the arrival phases of the flight, each type of activity that occurred, and each dimension of workload. Becoming consciously aware of the various aspects of workload requires training. Figure 11.6 provides descriptions of workload function and factor combinations that each pilot is asked to evaluate. A copy of this matrix is reviewed before each flight and is available with the questionnaire at the end of each flight leg. At the end of the questionnaire there is space for comments the pilot may have concerning any aspect of the questionnaire or the flight. After completion of each flight sequence an analyst reviews the completed questionnaire with the pilot and solicits more detailed information about any unusual events or any particularly high or low workload experiences.

The bottom section in Figure 11.12 shows the part of the questionnaire where the pilot specifies the reference airplane used in his or her evaluation of the test airplane. Currency in the reference airplane is established by indicating whether the pilot has flown the reference airplane within the preceding 90 days. The identification of a reference airplane by the evaluation pilot serves to anchor the pilot's ratings and comments. It also helps to temper any biases a pilot may have for or against an individual design or design feature.

WORKLOAD FACTORS

DEPARTURE (ARRIVAL) WORKLOAD FUNCTIONS	Mental Effort	Physical Difficulty	Time Required	Understanding of Horizontal Position
1 Navigation	Compare the degree of MENTAL EFFORT required by the navigation function of this airplane during Departure (Arrival).	Compare the PHYSICAL DIFFICULTY of operating the navigation system during Departure (Arrival)	Compare the AMOUNT OF TIME devoted to navigation during Departure (Arrival).	Compare the influence of the navigation system on your understanding of the HORIZONTAL POSITIONING of the airplane during Departure (Arrival).
2 FMS Operation and Monitoring	Compare the degree of MENTAL EFFORT necessary to Monitor and Operate the FMS with that required to accomplish similar functions in the rel. airplane during Departure (Arrival).	Compare the PHYSICAL DIFFICULITY of operationg the FMS with that required to accomplish similar functions in the rel. airplane during Departure (Arrival)	Compare the TIME REQUIRED to operate the FMS with that required to accomplish similar functions in the rel. airplane during Departure (Arrival).	— Blank —
3 Engine/Airplane Systems Operating and Monitoring	Compare the degree of MENTAL EFFORT necessary to Operate and Monitor the Engines and Airplane systems (other than FMS) during Departure (Arrival).	Compare the PHYSICAL DIFFICULTY to Operate and Monitor the Engines and Airplane systems (other than FMS) during Departure (Arrival).	— Blank —	— Blank —
4 Manual Flight Patch Control	Compare the degree of MENTAL EFFORT necessary to control flight path and speed during Departure (Arrival).	Compare the PHYSICAL DIFFICULTY of controlling flight path and speed during Departure (Arrival).	— Blank —	— Blank —
5 Communications	— Blank —	Compare the PHYSICAL DIFFICULTY of operating the communications system during Departure (Arrival).	TIME AVAILABLE	USEFULNESS OF INFORMATION
6 Command Decisions	Compare the degree of MENTAL EFFORT necessary to interpret the information available for decision making during Departure (Arrival)	— Blank —	Compare the TIME AVAILABLE for decision making during Departure (Arrival).	Compare the USEFULNESS of the information available for decision making during Departure (Arrival)
7 Collision Avoidance	— Blank —	— Blank —	Compare the TIME AVAILABLE to do visual scanning for collision avoidance during Departure (Arrival).	— Blank —

Figure 11.6 Description of Workload Evaluation Function and Factor Combinations. (Boeing, 1982)

BOEING
757/767

Pilot Subjective Evaluation

o Pilots are asked to provide an assessment of the 757/767 workload functions and factors, (FAR 25, Appendix D) experienced during flight crew operations.

o Detailed instructions are attached in a separate booklet.

Please fill-in the following information:

• Airplane Model ☐ 757 ☐ 767 Date of Flight ____ / ____ / ____ (Month) (Day) (Year)

• Airplane Number ________ Flight Number ________ Test Number ________

• Questionnaire Was Completed: Date ____ / ____ / ____ (Month) (Day) (Year) Time ________ (Local)

• Pilot's Name ______________________

• Flight Crew Assignment This Flight: ☐ Captain ☐ First Officer

• Organization
  - Boeing: ☐ Flight Test Boeing: ☐ Other
  - FAA: ☐ Flight Test FAA: ☐ Other
  - Other: ______________________

• Reference Airplane: Please indicate which single airplane you are using as a reference (check one)

☐ 737	☐ 737(SP177)	☐ DC-9	☐ DC-9-80	☐ L1011
☐ 727	☐ 747 ☐ 707	☐ DC-8	☐ DC-10	Other________

Have you flown your reference airplane (or an approved simulator for that airplane) in the last 90 days ☐ Yes ☐ No

A representative of the 757/767 Flight Deck Integration (B-8765) will collect your completed questionnaire. For additional information contact D.M. Fadden

THANK YOU

Figure 11.7 Evaluator Background Data Sheet, Pilot Subjective Evaluation Questionnaire. (Boeing, 1982)

Normal Operations: Departure

1.0 General Departure Information

1.1 Departure Data

(a) Departure Airport____________________ (b) Take-Off Time____________________(Local)

1.2 Flight Conditions for Departure

(a) Departure Airport Weather
- ☐ Less than 400 ft and 1 mile
- ☐ 400 ft and 1 mile to 1000 ft and 3 miles
- ☐ Better than 1000 ft and 3 miles

(b) Precipitation at Departure Airport
- ☐ None
- ☐ Light
- ☐ Moderate
- ☐ Heavy

(c) Meteorological Conditions Aloft During Departure
- ☐ VMC
- ☐ IMC
- ☐ Mixed

(d) Turbulence During Departure
- ☐ None
- ☐ Light
- ☐ Moderate
- ☐ Severe

(e) Other Significant Weather____

1.3 ATC Data Associated with Departure

(a) ATC Procedures Used During Departure
- ☐ VFR
- ☐ IFR: Vectoring Only
- ☐ IFR: Assigned Route
- ☐ IFR: Vectoring + Assigned Route

(b) Did you enter an amended route into the FMC/CDU after takeoff?
- ☐ Yes ☐ No

(c) Level of Interaction with ATC During Departure
- ☐ Low
- ☐ Moderate
- ☐ High

(d) Number of Altitude Clearance Changes During Departure
- ☐ 1 - 2
- ☐ 3 - 4
- ☐ 5 or more
- ☐ None

1.4 FMS Modes Used During Departure (Check applicable modes.)

(a) EHSI Use

	WXR	
☐ MAP	☐ Yes	☐ No
☐ VOR/ILS	☐ Yes	☐ No
☐ Both	☐ Yes	☐ No
☐ Neither		

(b) Autopilot Use
- ☐ CWS
- ☐ CMD — { ☐ LNAV ☐ VNAV ☐ Other }
- ☐ Not Used

(c) Flight Director Use
- ☐ Full-Time
- ☐ Part-Time
- ☐ Not Used

(d) A/T Use
- ☐ Full-Time
- ☐ Part-Time
- ☐ Not Used

Figure 11.8 Departure Information Data Sheet, Pilot Subjective Evaluation Questionnaire. (Boeing, 1982)

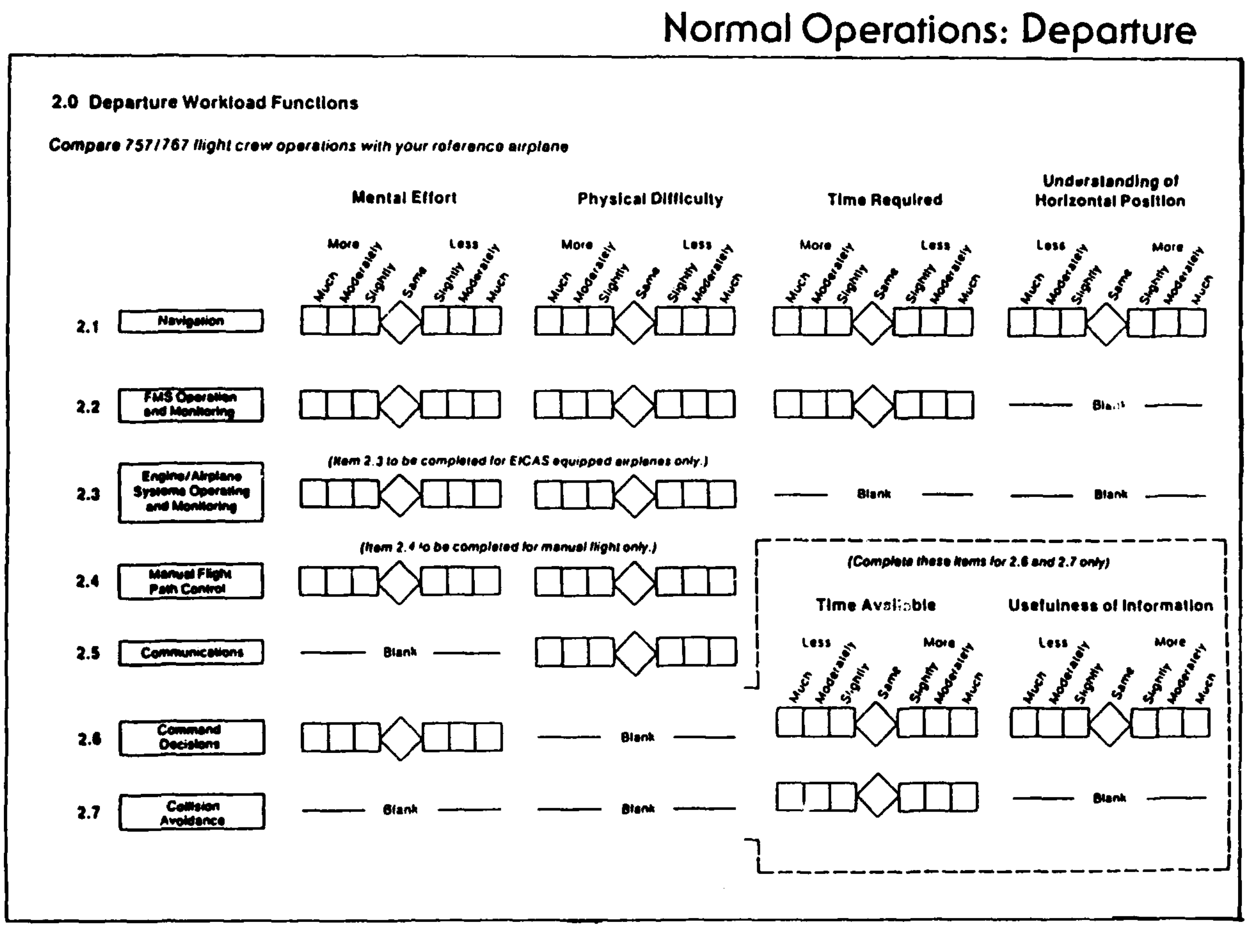


Figure 11.9 Departure Workload Rating Sheet, Pilot Subjective Evaluation Questionnaire. (Boeing, 1982)

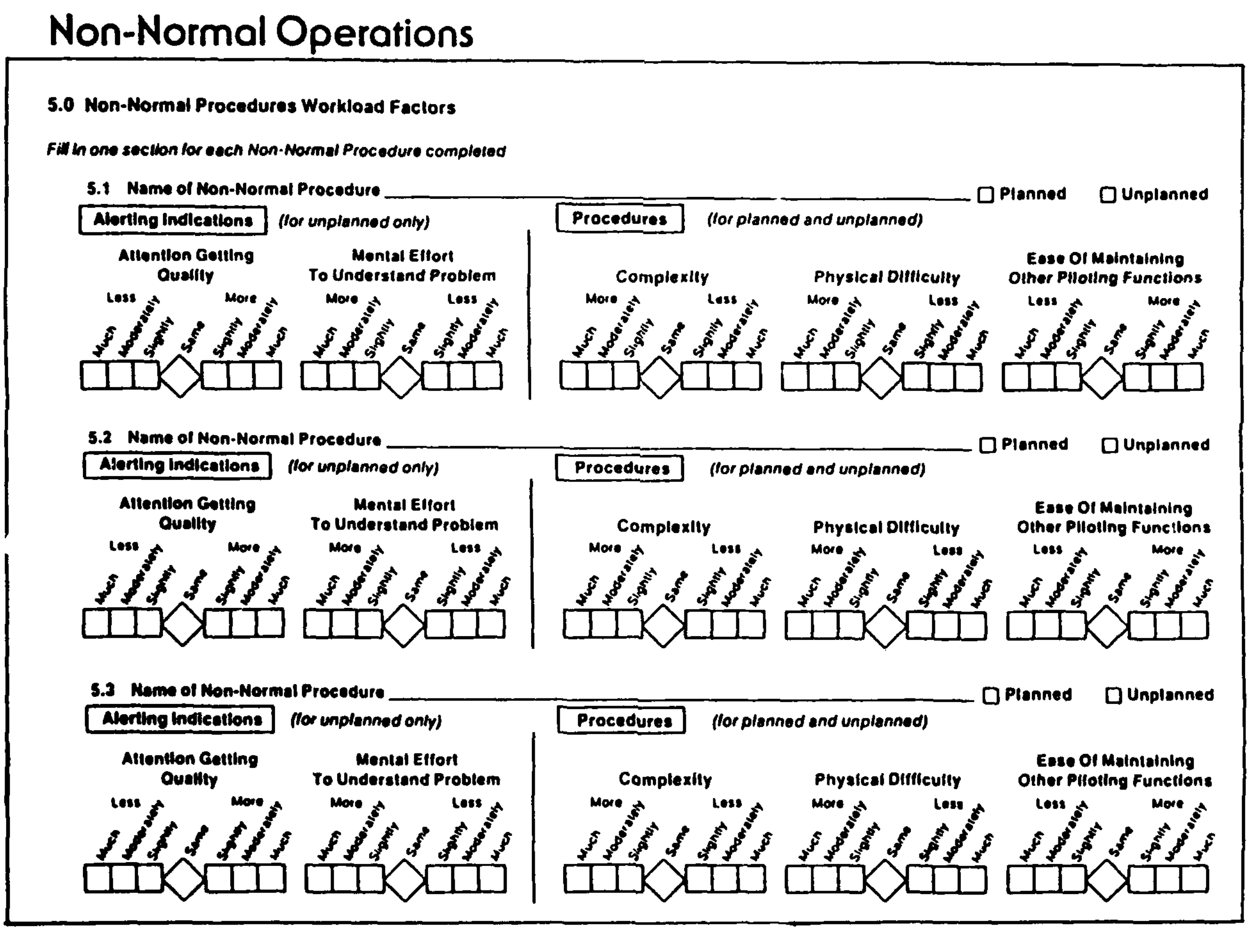


Figure 11.10 Nonnormal Operations Workload Rating Sheet, Pilot Subjective Evaluation Questionnaire. (Boeing, 1982)

Figure 11.11 is an enlargement of the rating boxes used in the Boeing questionnaire. This particular rating is for the Physical Difficulty of a function. The response boxes are arranged with increasing "goodness" to the right. The leftmost box indicates the greatest workload in comparison with the other airplane, while the rightmost box indicates the least workload. For example, if the pilot must exert much more physical effort to perform the task in question when flying the 757 than when flying the reference airplane, the pilot simply checks the box farthest to the left. The condition where workload is essentially the same for both the 757 and the reference airplane is indicated by the diamond in the center of the workload scale. For some of the workload or evaluation factors in the questionnaire, the labels "More" and "Less" are reversed from the sense in this figure. Consistency is retained in that goodness continues to increase to the right in all cases.

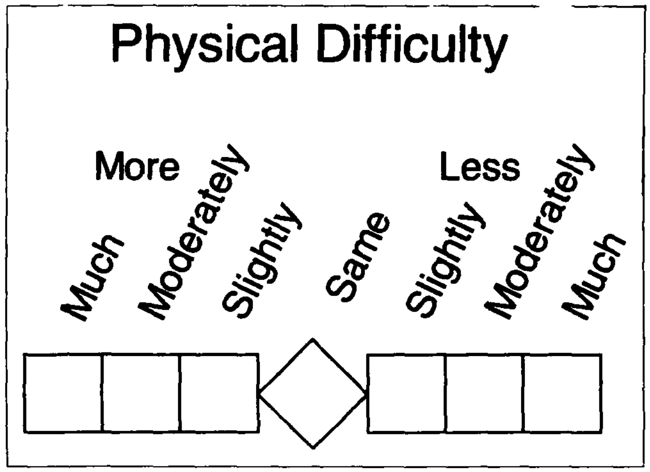


Figure 11.11 Rating Boxes Used in the Boeing Pilot Subjective Evaluation Questionnaire. (Boeing, 1982)

The evaluation forms developed by Boeing for the 757/767 focus on departure and arrival activities and nonnormal procedures where workload is highest and most variable. The questionnaire was originally drafted as eighteen pages of text-based questions. In this form, it was explicit enough to be used without training; however, after using the form several times, many evaluators objected to having to read so much material. Along with the objections, the rate of inconsistent answers on the questionnaire increased. With the help cf a consultant skilled in questionnaire development, the text format was changed to a graphical one reducing the page count by two-thirds.

The basic portion of the questionnaire asks for ratings for mental effort, physical difficulty, and time for each of the significant workload functions (see Figure 11.6). Where both equipment and procedures on the new airplane are conceptually identical to those on current airplanes, the rating request is deleted. The time required rating presents a particular problem for the evaluator and the analyst. Studies by Sandra Hart at NASA-Ames have shown that people are poor judges of time when they are involved in highly skilled tasks. To make matters worse, the pilot is not likely to recognize when his time estimates are good and when they are not. In deciding when to ask for time estimates, we

gave significant weight to those tasks that have a high conscious activity content. These choices were then subjected to review during the simulator validation of the questionnaire.

While the core content of the form applies equally well to any commercial airplane, the unique features of any new model might warrant special consideration. For example, the 767 included a CRT map display and a full-time flight management system. There was some concern that these devices would add workload. To understand better the total impact of these devices, two questions were added to the questionnaire dealing with the information supplied by these systems. These questions were integrated into the questionnaire and appear in the far right column of Figure 11.6. They are titled, "Understanding of Horizontal Position" and "Usefulness of Information."
The nonnormal operations portion of the questionnaire (Figure 11.10) provides additional workload information about equipment failures or abnormal flight conditions. These events always involve two aspects: recognition of the event or condition and accomplishment of any special handling required to restore normal operations. The questionnaire asks for two ratings regarding the alerting indications and three ratings about the nonnormal procedure itself. In this case, the mental effort rating is titled "Complexity" and the time required rating is titled "Ease of Maintaining Other Piloting Functions." These enhancements resulted from discussions with pilots who found these titles easier to relate to the specific events of a nonnormal procedure.

During flight test operations, there is the possibility that actual equipment failures or nonnormal flight environments will occur. Even though these unplanned events are not specified in the test plan, they are included in the nonnormal portion of the questionnaire process. Where possible, simulated inflight faults are introduced in a way that will produce the appropriate alerting and recognition indications to the pilot. These events are also treated as unplanned on the questionnaire, since they appear to be unplanned from the viewpoint of the evaluation pilot. Safety concerns limit the failure event realism that can be simulated inflight. Where such concerns come into play, the alerting indications will be missing or incorrect. However, the procedure portion of the questionnaire is still valid and useful.

Normally, the questionnaire is completed by the evaluation pilots for both the departure and arrival phases of the current flight leg immediately after landing and before any discussion takes place. On occasion, the departure sheet can be completed once the aircraft reaches cruise altitude; however, the requirements of the test program generally place heavy demands on the pilots while airborne. The post flight debriefing involving the evaluation pilots and a human performance analyst is an important element in the total process. Through this

debriefing, additional material is collected giving a complete understanding of the events that each pilot felt were significant contributors to the ratings.

Initial validation of the questionnaire was done in the simulator using a variety of test and training pilots. This was followed by trial use during developmental flight testing of the 767. The questionnaire was used during the minimum crew size proving flights for the 767 and later for the 757. With appropriate adjustments it was also used during the minimum crew size proving flights for the 747-400.

The pilot subjective evaluation process provides nonscalar ratings for specific workload functions; as such, the ratings are not amenable to summary combination. Various people on all sides of aircraft certification would like to have a single number or rating to characterize the airplane. The present state of human performance knowledge does not provide a simple and meaningful basis for combining the PSE ratings. Future research may provide new insights that will make such a combination meaningful. For the time being it will continue to be necessary to repeat the explanation of why arithmetical combinations of the ratings are not meaningful.

One final issue surrounds the use of subjective ratings as a part of aircraft certification. Who should do the evaluations? Clearly pilots from the responsible regulatory agency must be involved. The manufacturer has a central role since it is the manufacturer who is offering the aircraft for certification, and it is the manufacturer who bears total responsibility for the aircraft until it is delivered to the final customer. Test pilots from the manufacturer and the regulatory agency are the best trained evaluators. They know the airplane well through exposure during the development program and have seen it perform through many tests, some of which exceeded the flight envelope boundaries of line operation. The regulatory agency pilots who are responsible for training and overseeing line operations have an insight into the full variety of airline operations that exceed the experience of most line pilots who fly with a single airline. These two groups should constitute the bulk of the evaluation pilot pool.

The use of line pilots in the certification program has been suggested on various occasions. We believe that line pilots are better used early in the development program and for simulator tests of new functions and features where airplane performance can be measured along with the pilot's opinion and differences can be resolved through discussion and further testing. In any case, if line pilots were to be directly involved in the evaluation, it is likely that significant changes would have to be made to the overall test program to compensate for the lack of evaluator training and to assure sufficient

standardization of this subjective process that the results obtained can be interpreted. Such steps would be necessary to protect the regulatory agency, the manufacturer, and the line pilot himself from errors of commission and omission in the evaluations. In recent certification programs, the FAA has asked a few retired industry pilots to consult during the crew size flight testing. In this way, the ultimate authority for the certification decision has remained with the FAA while an additional source of information and review has been made available. This program appears to have worked satisfactorily for all parties.

Certification Considerations

Early Requirements Determination

One of the driving issues in airplane manufacturing today is reshaping the structure of the design-build cycle in ways that will improve the efficiency of the process so that the right airplane is designed and the airplane is built right the first time. The factors that make this effort mandatory are deeply rooted in the commercial aviation marketplace. Cost is a major factor but, so too, is time-to-market. These changes cannot be accomplished at the expense of safety. At the same time, safety cannot be used as an excuse for not finding ways to satisfy market demands. Many people believe that the needed process changes mandate both earlier and more complete determination of requirements. In this context, the word "early" means that requirements are understood and documented before the airplane is built. This places a significant burden on FAA certification personnel. The certification system itself is designed to place primary emphasis on near term certification programs. Furthermore, there is a strong tradition of withholding judgment until the completed product is available. Finding ways to uncover the majority of concerns while the design is still on paper, and yet maintain the objectivity necessary for the final approval, will be challenging indeed.

Mandatory Indicators and Displays in Integrated Cockpits

Mandatory displays, particularly those defined explicitly in terms of their format, are another problem. Most mandatory displays are the result of previous accidents or highly focussed public concerns. Required displays reflect aircraft operations and the pilot interface understanding that exists at the time they are first developed. Over time, both aircraft operations and the pilot interface understanding evolve and, as they do, the displays, indicators, and procedures that characterize the flight deck change as well. Eventually the gap between the current displays and the mandatory displays becomes great enough that there is concern that effective pilot performance will be retained.

A good example of this difficulty is the handling of indications alerting the flight crew to equipment failures or abnormal operational conditions. By the mid-1960s, the number of independent indicators had grown to the point where people within the FAA, the airlines, and the manufacturers were concerned. An FAA-sponsored study, done jointly by Boeing and Douglas, developed and validated the concept of a centralized caution and warning system. The concept has been widely embraced and is implemented in the 767 and subsequent airplanes. Certifying the system on the 767 required an equivalent safety ruling from the FAA. Even today, if a manufacturer abides strictly by the rules, the resulting flight deck will contain an array of dedicated red and yellow lights and a multitude of alerting sounds. No one questions the intent of those who established the initial mandatory display requirements. The concern is that conditions have changed. Our collective understanding of human performance has improved and the technology available to satisfy operational needs has changed. It is time to recast some of the very specific design rules with the performance they are meant to achieve.

Airline Differences

Another certification consideration that poses a problem for both the FAA and manufacturers is airline difference. Airlines are different. They have different fleet mixes. They operate in different regions. They have different crewing policies. They have different strategies for achieving operating standardization. These differences exist among domestic airlines and even more among foreign carriers. It is important that these differences be understood and accommodated in the certification process. The apparent efficiency that some believe would follow from enforced flight deck standardization may be an illusion. There certainly are standardized features that benefit the entire industry; e.g., direction of movement of primary controls, general layout of the primary instrument panel, and minimum instruments for IFR flight. However, each feature should be judged on its own merits before concluding that standardization is the appropriate path. Even when standardization is chosen, the choice should be re-evaluated at regular interval to determine if it is still the appropriate action.

Equipment standardization and operations standardization are not synonymous. If fundamental airplane or equipment performance differences force operations to be different, standardizing equipment will not achieve operation standardization. It may, in fact, interfere with safe and efficient operations by creating the illusion of consistency where it does not, and should not, exist. The standardization debate would be better served by addressing the fundamental principles that underlie effective human performance. This approach has

significantly greater potential to combat the consequences of human error, though it is much more difficult to accomplish.

Coping with Pilot Error

Error Types

Since accidents are the most serious consequence of human error, significant time and effort are spent evaluating accident and near accident situations. A consistent finding is that several errors occurred before the accident was unavoidable. Studying crew-related accidents helps identify possible error sequences and patterns and may lead to an understanding of the factors that kept the crew from recognizing the seriousness of the situation until it was too late. The ultimate goal is preventing errors that cause accidents. Helping the pilot break the error chain before an accident is inevitable is one of the ways of achieving the goal. Errors that result from clearly understood events or circumstances can be handled more directly by the designer and the pilot than those resulting from unknown conditions. Because of the difference in management and coping strategies, we find it convenient to classify errors as either systematic or random, respectively.

Through careful design, systematic errors can be reduced to a very small number and the pilot can be trained to recognize and deal with those systematic errors that cannot be eliminated. Minimizing systematic errors involves careful attention to human factors data and rigorous attention to the design development process. The unspecific nature of random errors makes their elimination more problematical. Human performance research will, over time, uncover the knowledge that converts random errors into systematic errors that then can be eliminated. Meanwhile, design strategies, such as system simplification and the minimization of time critical procedures, can reduce the opportunities for random error. In the end, however, ensuring that the pilot can detect that an error has occurred and can do something about it, is the best means of preventing the error from compounding into a more serious situation. This is the essence of error tolerance--detection and effective action.

Error Tolerant Design

If the pilot is to cope with the error, the pilot must first detect it or have it pointed out. Direct feedback of pilot actions is an obvious way to helping the pilot to detect an error. In some circumstances, direct feedback is not practical or simply cannot be done. Under these circumstances, enhancing situational awareness for the pilot can provide a framework within which certain errors can be detected. Providing redundant, dissimilar cues is another useful error

detection technique, particularly where the consequences of an error would be costly. This technique is particularly valuable where the human tendency to perceive what is expected, even in the presence of contrary cues, is the root cause of the error. Of course, detection of certain errors can be done by systems on the airplane and their existence announced to the pilot. This widely used technique can be highly effective where response time requirements are compatible with human capabilities.

Once an error has been detected, the pilot must be able to react in a way that reduces the likelihood of the error sequence continuing. Often the reaction will be to accomplish some physical action. Under other circumstances the appropriate reaction may be a change in planning or strategy for the remainder of the flight. Recognizing the full range of possible responses is the key step in ensuring that the pilot is provided with the appropriate controls, information, knowledge, and skills to react effectively.

One of the most difficult aspects of pilot error is recognizing what errors are most likely. It is nearly impossible for one human being to imagine how another human being could understand and interpret the same circumstances differently, yet evidence abounds that such is the case. Add to this that pilots vary considerably in their decision-making styles and it is evident that understanding error is a team effort. Collective wisdom is consistently one of the more effective means of seeking out possible error patterns and their causes. For collective wisdom to work, it must be nonjudgmental with an emphasis on understanding as many ways of interpreting the display or control device as possible. There are no wrong answers, except to believe that one interpretation is correct and the others are wrong. The goal must be to help all pilots catch their mistakes.

Pilot error can be triggered by unrecognized and subtle mismatches between the information that is presented and the tasks that information is meant to support. It is easy for the pilot to assume that if the information presented is the same, then the associated tasks must be the same as well. Conditions where identical indications are used to support different tasks are an invitation to error. To make matters worse, error detection by the pilot under these circumstances is particularly difficult. Making the design error-tolerant means that the possibility of this error is acted upon during design. If the assumption that the tasks are the same is false, the simplest design solution is selection of different display formats, indications, or controls.

As an example, the hydraulic systems of the 757 and 767 are slightly different operationally, because of different load assignments to the individual hydraulic systems. Slight differences in system management and post-failure planning

result from this difference. Because of the task difference, the hydraulic system control panels on the two airplanes are intentionally different. Even though the same number of control devices is required on both airplanes, the types of switches and the physical layout of the panel are different.

Boredom, fatigue, and time-of-day are among the factors that influence pilot attentiveness. Their effects will normally vary during a single flight. Given these facts, it is obvious that the pilot cannot be at maximum attentiveness all the time. The design of nonnormal procedures can be made error-tolerant by ensuring that the pilot has extra time to recognize and respond to situations that, from his perspective, are new or unexpected. Once alerted to the possibility of a problem or unusual condition, virtually all pilots can achieve significantly increased levels of attention within a short time. This heightened attention can then be sustained, if the circumstances warrant, for much longer than it took to reach the heightened attention initially.

Future Workload Issues

In the future, crew workload will be influenced strongly by the strategy used to prioritize flight deck information. Pilots are expected to look at, and be aware of, an ever increasing array of information. Human beings can be exceptionally versatile at handling large quantities of information. However, the time pressure of flight can lead to impromptu prioritizing strategies that may not be well suited to the actual circumstances. While certifying an individual system, the composite effect of that system on the total flight deck information load may not be evident. Yet the overall flight deck information management issues can only be addressed by managing the contribution of each system. This means that everyone involved in development and certification of specific equipment or systems must share responsibility for the impact of those systems on the overall pilot-airplane interface effectiveness.

A related issue is the potential for information overload that could follow the addition of a general purpose data link capability to the airplane. Conceptually, such systems could allow the nearly limitless information sources stored in ground-based computers to be available in the cockpit. The potential for good is great but so is the potential for excessive information management workload. The knowledge and the tools are available to ensure that realistic consideration is made of the pilot's human capabilities and limitations. The question is, how will we, as an industry, use this information to ensure that new data sources are managed in a manner that improves the effectiveness of the pilot and protects the aviation system from new human error risks.

In the future, it is conceivable that the basic reliability of some of the control and display equipment will approach the lifetime of the airplane. This implies that many crews will go through their entire careers without seeing certain first failure conditions on the actual airplane. While this will reduce nonnormal workload, it presents some interesting challenges for selecting appropriate fault management strategies and training. Certainly the strategies of today, based on memorized or highly practiced procedures, will be inefficient and may not be effective. The assistance of computer-based expert systems may be desirable. Alternatively, it may be better to create designs that ensure the pilot will have time to develop a suitable response by applying his knowledge of the system or event.

A final issue concerns the increasing performance demands placed on pilots and systems by the increasing need for aviation system efficiency. Many of the improvements in efficiency are likely to result from better matching of: the information available to the pilot, the procedures established for the various tasks, and the training the pilot receives. To avoid any unnecessary increase in pilot workload, coordination of these improvements will require more communication and understanding among all the organizations and agencies involved. It will take foresight and initiative to weld the traditionally independent domains of aviation equipment and operations into a team that enables the American aviation system to remain the best in the world.

Chapter 12

Human Factors Testing and Evaluation

by Kim M. Cardosi, Ph.D., Volpe Center

Introduction

Many different types of questions are best answered with the results of a human factors test. Some of the most common human factors questions include:

- Which of two or more proposed designs (of displays, controls, training programs, etc.) is best from a human factors standpoint?
- What performance benefits are achieved from a specified design change?

- What performance benefits are achieved from a specified design change?
- Is a design of a new system or subsystem viable from a human factors perspective?
- What changes, if any, need to be made to a prototype system to minimize operator error?
- Is a proposed training program (e.g., for new equipment) adequate?
- How long will it take for an operator to perform a task, or part of a task, with a new system?

Human factors specialists, working with operations specialists, can often anticipate human factors problems by examining specifications documents, proposed designs, and prototypes of new systems and subsystems. Still, human factors tests are often required to identify problems that are not self-evident or to be able to quantify the impact of new systems on line operations. Formal evaluations are always needed to ensure that the new system or procedure is ready for implementation.

This chapter will address the following questions.

- When is a human factors test warranted?
- How is operator performance measured and what factors can affect these measures?
- What method of testing should be employed?
- How should test results be analyzed and interpreted?

Understanding the principles and philosophy behind human factors testing is useful even to people who never conduct human factors tests because it helps operations specialists critique tests conducted by manufacturers, universities or industrial labs and determine the validity of their conclusions.

When Is a Human Factors Test Warranted?

It is not always easy to predict all of the ways in which an operator will use or misuse a new system or a new component of an existing system. Nor is it always evident what types of errors that operators are likely to make. One example of a faulty display design that should have never made it to

implementation is the case of a major air carrier that wanted to give the flight attendants a cue as to when sterile cockpit was in effect. The airline installed a small indicator light above the cockpit door that was to be illuminated when sterile cockpit was in effect. Problems arose because the light that was chosen was green. In most cultures, the color green is **not** associated with "stop" or "no admittance." The lights had to be changed to red, at no small expense to the airline. In that case, a human factors test was not needed to predict the problems that were experienced by the airline; it is common knowledge how a green light is likely to be interpreted by a crewmember. However, most questions about training, displays, controls, and how the operators may use or abuse them are much more complex and require controlled testing to be answered effectively.

The findings of basic research, such as information about our sensory and cognitive capabilities and limitations, can steer us away from what is known to be troublesome and can help us to identify desirable design options. However, each specific application of a technology, training program, or procedure should be evaluated under the same or similar conditions as it will be used, by the same type of operator that will be using it, and while the operators are performing the same types of tasks that actual operations require.

When a human factors evaluation of a system or subsystem is warranted, it should be designed by both a human factors specialist and an operations specialist. Operations specialists are intimately familiar with the operational environment (e.g., a specific cockpit or ATC facility). They represent the potential users and are usually operators (e.g., pilots or controllers) themselves. As long as they are operationally current (i.e., knowledgeable of current issues, procedures, and practices), they are the most appropriate source for information on user preferences and suggestions for symbology, terminology, display layout, etc. However, even the most experienced users should *not* be solely responsible for the user-machine interface. In fact, many years of experience can occasionally be a liability in making such decisions, since the skills and knowledge that develop with extensive experience can often compensate for design flaws that may then remain unnoticed. For these and other reasons, it is important for operations specialists to work with human factors specialists in the planning and conduct of a human factors test. Human factors specialists are intimately familiar with the capabilities and limitations of the human system, testing methods, and appropriate data analysis techniques. They can point to potential problems that operational specialists might otherwise overlook. While working together, the two specialists can predict problems and head them off before they occur in actual operations. Together, they are best equipped to decide exactly what needs to be tested and how it should be tested.

How Is Human Performance Measured?

Measures of human performance can be subjective or objective. Subjective measures use responses that are measured in terms of the person's own units. Such measures can be influenced by the individual's expectations and motivations. An example of a subjective measure of workload is a pilot's opinion as to how difficult a task is. What constitutes a high workload situation for one person may not be considered high workload for another person. Subjective measures are used whenever objective measures either aren't available, or aren't appropriate. They're also used to complement objective measures.

Objective measures of human performance use units that are clearly defined, such as seconds, or percent errors, heart beats per minute, blood pressure, etc. The most commonly used objective measures of performance are response accuracy and response time. Response time measures the time required for a person to perform a specific task, or component of a task. Response accuracy measures the percentage of errors made while completing the task or the precision with which a specific task is accomplished (such as flying a predetermined route, as measured by cross-track error).

When measuring only response accuracy, it is possible to obtain insignificant results due to either a ceiling effect or a floor effect. That is, the response being measured may be so skilled, (e.g., a baseline of 95 percent accuracy) that any manipulated factor is not likely to have an observable effect. This is called a ceiling effect. Conversely, initial performance may be so poor that any manipulation will not have a measurable effect. The tests may not be sensitive enough to measure an effect beyond this very high or very low baseline.

Generally, if baseline performance on the measured task is extremely accurate, and it is not desirable to induce more errors by manipulating other factors (e.g., workload), then response time is generally a more sensitive measure than error rates. Differences in the response times may be observable even when the differences in response accuracy are not.

Components of Response Time

While response time appears to be a simple measure of human performance, it is actually quite complex. Response times have several components and each of these components can be affected by many different factors. These factors must be considered in any human factors test so that the controls necessary for confident interpretation of the data can be employed.

A complex response, such as one to a cockpit warning system, may be broken down into four components: detection time, time to identify and interpret the message, decision time, and time to initiate (or complete) the appropriate response. When a warning signal appears, for example, the first component of the required response is to detect the presence of that signal (i.e., the warning message), that is, to notice that it is there. The second component of the response is the interpretation of the message. The operator needs to identify the message. For example, is it TCAS, or GPWS? While this stage may sound simplistic, the task becomes more difficult as the number of alarms and warnings increases. After deciding which message it is, the next response component required is to decide what physical action, if any, (e.g., a climb or turn) is required. Then, and only then, can a physical response be initiated.

Results of a series of flight simulation studies indicate that, with an executive system, (that is, one that requires immediate action) it will take approximately two to three seconds to detect that the message is there, five to six seconds to decide what to do about it, and one to two seconds to initiate a response (Boucek, White, Smith, and Kraus, 1982; Boucek, Po-Chedley, Berson, Hanson, Leffler, and White, 1981; Boucek, Erickson, Berson, Hanson, Leffler, Po-Chedley, 1980; see also Berson, Po-Chedley, Boucek, Hanson, Leffler, and Wasson, 1981). This leads to a total of eight to eleven seconds that should be allotted for a pilot to respond.

The most stable of these components, that is, the one that has the most predictable duration, is the initiation of the physical response. Since the decision as to what action is required has already been made, the initiation of the response constitutes the smallest component of response time. The time required to complete the response will, of course, depend upon the task.

Factors Affecting Human Performance

There are many factors that are known to affect human performance, and hence, response time. Some of these factors are characteristics of the stimulus, that is, of the visual or auditory display. Others are characteristics of the operator, such as, previous experience, skill, fatigue, etc. Still others are characteristics of the test or operational environment, such as workload, consequences of errors, etc. Each of these factors needs to be considered from the test design to the interpretation of the results and controlled as much as possible during a test.

Stimulus Factors

Factors that influence detection of visual signals include location in the visual field, and presentation format (e.g., blinking vs. steady text, brightness, etc.). (See Chapters 1 and 2 of this text.) Response time will be faster if the signal is presented in the center of the visual field, as opposed to out on the periphery. If it is presented in the periphery, but flickering, detection time will be faster than if it is in the periphery but steady. (This is one reason why a flickering display can be distracting.) Intensity is also an important factor. Within limits, a higher intensity stimulus will attract attention more efficiently than a less intense stimulus. In the visual domain, intensity translates into brightness (although other factors, such as contrast) are also critical. For auditory displays (e.g., a tone or spoken warning message), intensity translates into loudness, with frequency as a critical variable. The frequencies that are contained in the ambient noise must be considered in deciding which frequencies should be contained in the alert. The relative intensity of a message (tone or voice) must always be measured in the environment in which it will be used. A warning signal that sounds very loud on the bench may be inaudible in a 727 with the windshield wipers on. In fact, the original Traffic Alert and Collision Avoidance System (TCAS) voice alerts passed the bench test, but were found to be unusable in the cockpit (Boucek, personal communication).

Meaningfulness

Another factor that can affect how quickly a signal can be recognized and interpreted is how meaningful the signal is. Personally meaningful stimuli, such as one's own name, and culturally meaningful stimuli, such as the color red or a European siren (both of which are associated with danger) will attract attention more efficiently than other stimuli of equal intensity. One exception to this, however, is if one of these "meaningful" signals is presented repeatedly without accompanying important information (as with false alarms). In this case, it is not difficult to learn to ignore a signal that previously attracted attention efficiently.

Ease of Interpretation

Another factor that affects response time is how intuitive the meaning of the symbol is to the user. For example, one of the first TCAS prototypes used a red arrow to convey to the pilot the urgency of the alert (red) and the direction in which the pilot should fly. Even after training, some pilots felt that there could be instances in which pilots would be unsure as to whether a red arrow pointing up meant that they should climb or that the traffic was above them. The arrow was changed to a red or amber arc on the IVSI (Instantaneous Vertical Speed Indicator) with the instructions to the pilot to keep the IVSI

needle out of the lit (red or amber) band. This provided a more consistent coding between the urgency of the alert and the required action.

Expectations and Context

Expectations and context have a strong influence on response time. Responses to a stimulus that occurs very frequently, or one that we expect to occur, will be faster than to one that occurs once every month. However, expectations may also lead to inaccurate responses, when what is expected is not what occurs. In many situations, particularly ambiguous ones, we see what we expect to see and we hear what we expect to hear.

The following ASRS report (October, 1989) entitled "Something Blue" illustrates the power of expectation:

> "On a clear, hazy day with the sun at our backs we were being vectored for an approach...at 6000' MSL. Approach advised us of converging IFR traffic at 10 o'clock, 5000', NE bound. After several checks in that position I finally spotted him maybe 10 seconds before he passed beneath us... When I looked up again I saw the small cross-section and very bright landing light of a jet fighter at exactly 12:00 at very close range at our altitude... I overrode the autopilot and pushed the nose over sharply. As I was pulling back the thrust levers and cursing loudly, the "fighter" turned into a silver mylar balloon with a blue ribbon hanging from it! I could see what it was when it zipped just over our heads and the sunlight no longer reflected directly back in my eyes (the landing light). I was convinced it was a military fighter, complete with the usual trail of dark smoke coming out the back (the blue ribbon?)!
>
> Then -- I remembered the traffic directly below us! I pulled the nose up just as sharply as before. Fortunately, everyone was seated in the back, and there were no injuries or damage... Our total altitude deviation was no more than 200'."

In this case, the expectation or "set" to spot traffic led to a false identification of an object and, consequently, an inappropriate response to it.

Another good example of the powers of expectation is seen in the videotapes that Boeing made of their original TCAS simulation studies. In this study, the pilots had the traffic information display available to them and often tried to predict what TCAS was going to do. In one case with a crew of two experienced pilots, the pilot flying looked at the traffic alert (TA) display and said, "I think we'll have to go above these two guys" (meaning other aircraft). This set up the expectation for both crewmembers for a "climb" advisory. The

crew started to climb when they received their first TCAS message, "Don't climb." The pilot flying told the pilot not flying to call Air Traffic Control (ATC) and tell them what action they were taking. Without reservation, the pilot not flying called ATC, said that in response to a TCAS alert, they were climbing to avoid traffic. He also requested a block altitude. He then told the pilot flying that they were cleared to climb. Meanwhile, as the climb was being executed, "Descend" was repeated in the background over 25 times. Eventually, the pilot not flying said, "I think it's telling us to go down." The next thing that is heard on the tape is "[expletive], it changed, What a mess." Crash. (Boucek, personal communication).

Anyone could have made a similar mistake. It is human nature to assess a situation and form expectations. In support of the pilot's expectation, and perhaps because of it, he didn't hear the first syllable, which was "don't" - he heard the action word "climb". The idea was then cemented. It takes much more information to change an original thought than it does to induce a different original thought.

Practice

Another factor that affects response time is how practiced the response is. If the response is a highly-practiced one, then response times will be quicker than if it is a task that isn't performed very often.

User Confidence

Another important factor is trust in the system. This may, or may not, develop with exposure to the system. Response time will increase with the time required to evaluate the validity of the advisory. Confidence in the system and a willingness to follow it automatically will result in shorter response times.

Number of Response Alternatives

Another factor that influences the decision component of response time is the number of response alternatives. In Ground Proximity Warning System (GPWS) for example, once you decide to respond, there is only one possible response: to climb. In TCAS II there are two response alternatives: to climb, or to descend. With TCAS III, there are at least four alternatives: climb, descend, turn right, or turn left. Studies have shown that the response time increases with the number of response alternatives (see Boff and Lincoln, 1988 p. 1862 for a review).

"Real World" Data on Pilot Response Time

It is difficult, if not impossible, to fully simulate the operational environment in even the most sophisticated simulation facilities. For this reason, data on pilot response time that is obtained unobtrusively from observational studies of "real world" events is extremely valuable (but rarely available). There are at least two such studies of pilot behavior. One examines pilot response times to GPWS and the other to time-critical ATC communications.

Ground Proximity Warning System (GPWS)

Several large overseas international airlines measured pilot response times to a time-critical GPWS warning - mode 2 "Terrain-Terrain" (which indicates high speed flight toward rising terrain). This information was collected during actual flights and indicated that the pilot response times ranged from 1.2 to 13 seconds with an average of 5.4 seconds (Flight Safety Foundation, Accident Prevention Bulletin, January 1986). No other statistical information on pilot response time (such as how many data points were included in this sample or the response time at the 90th percentile) was reported. It is also interesting to note from this study that even though the Boeing recommendation was an initial pull-up of 15 degrees, and the Douglas recommendation was an initial pull-up of 20 degrees, the average pull-up observed was 8.5 degrees with a rotation rate of 1.4 degrees per second. This may be inadequate in many terrain encounters.

Air Traffic Control (ATC)

In an analysis of pilot response time to time-critical ATC transmissions in an en route environment, Cardosi and Boole (1991) analyzed 46 hours of controller to pilot communications from three Air Route Traffic Control Centers (ARTCCs). In these 46 hours of voice tapes, 80 communications from controllers to pilots were found to contain time-critical messages, such as maneuvers required for traffic avoidance, or maneuvers followed by words expressing urgency (e.g.,"now" or "immediately"). The pilots' verbal response times, as measured from the end of the controller's transmission to the beginning of the pilot's acknowledgement, ranged from one to 31 seconds with a mean (i.e, average) of three seconds (standard deviation = 5). The 90th percentile was 13 seconds. This means that we would expect most (90%) of pilot responses to be initiated within 13 seconds. The average response time, as measured from the end of the controller's transmission to the end of the pilot's initial transmission (even if it was only a "say again") was six seconds.

To measure response time from a systems approach, Cardosi and Boole examined the total time required for successful transmission of a time-critical

message. This was measured from the beginning of the controller's transmission to the end of the pilot's correct acknowledgement (and included "say agains" and other requests for repeats). This total time ranged from four to 40 seconds and averaged 10 seconds. Ninety percent of the transmissions were successfully completed within 17 seconds. Interestingly, times required to complete similar, but not time-critical transmissions, such as turns issued by controllers for reasons other than traffic avoidance, were very similar. The time required for successful transmission of such calls ranged from four to 52 seconds with a mean of 10 seconds.

Finally, it is interesting to note that many pilots' (and controllers') perception is that a pilot's responses to GPWS and to time-critical calls is immediate. While this is largely true, analysis of the data shows that even the immediate takes time.

What Method of Testing Should Be Used?

The testing method of choice depends on the specific problem or question under investigation and the available resources. Most importantly, the method must be appropriate to the issue. For example, one would not consider a questionnaire for measuring the time required to complete a small task, nor would one collect data on pilot eye movements by asking the pilots where and when they moved their eyes. Another necessary consideration is the amount and type of testing resources available. Often, the most desirable type of test is too expensive and many compromises are necessary. The implications of these compromises need to be recognized as do their implications for the interpretation of the test results.

Field Observations

One evaluation technique that is often used is field observation. This includes any over-the-shoulder evaluations, such as sitting behind the pilot and observing a specific pilot activity or sitting behind a controller team and observing their interactions. One advantage to this method is that it allows investigators to make observations in the most natural setting possible. It can increase our understanding of the nature of processes and problems in the work environment. Specifically, valuable insights can be gained as to where problems might occur with a specific system or procedure and why they might occur.

One task in which field observations are helpful is in trying to determine the information or cues that people use in performing a task. We, as humans, are rarely aware of all of the information that we use in performing a task. This is illustrated in a "problem" that Boeing Commercial Airplanes once had with one

of their engineering simulators. After flying the simulator, one pilot reported that, "It felt right last week, but it just doesn't feel right this week." The mechanics examined everything that could possibly affect the handling qualities of the simulator. They took much of it apart and put it back together. They fine-tuned a few things, but made no substantive changes. The pilot flew the simulator again, but again reported that it still didn't "feel right." It seemed a little better, but it just wasn't right. Someone finally realized that the engine noise had inadvertently been turned off. The engine noise was turned back on and suddenly, the simulator once again "handled" like the aircraft (Fadden, personal communication).

While field observations are often useful as initial investigations into a problem, their limitations often preclude objective conclusions. Their findings may be more subjective than objective, are dependent on the conditions under which the observations were made and can actually be affected by the observation process itself.

One factor that affects the reliability of findings based on field observations is the number of observations made. For example, a conclusion based on 10 test flights is going to be more reliable (i.e., more repeatable) than one based on three flights. Furthermore, the findings based on field observations are condition-dependent. That is, the findings must be qualified with respect to the specific conditions under which the observations were made. For example, if you observed five test flights and they all happened to be in good weather, with no malfunctions, et cetera, you may have observed only low or moderate workload flights. Any findings based on these flights can not then be generalized to situations involving high workload.

Another, and more subtle, consideration is that the very process of observation can alter what is being observed. An observer's activities, or even his or her mere presence, can affect performance. For example, depending on who the observer is (and their stated or implied mission), a flight crew may change their behavior. They may, for example, become more conscientious (e.g., about checklists). It is easy to envision how different observers (e.g., a university researcher, an air traffic controller, or an FAA inspector) might observe slightly different behaviors exhibited by the same crew, all of which may be different from what occurs when no observer is present.

Another possibility is that the observer's presence might make a crewmember nervous and induce a classic case of "checkitis". In this case, performance would be poorer than when no observer is present. Observers, or their questions, may also be distracting and this may adversely affect performance.

Questionnaires

Questionnaires are important research tools that allow investigators to collect information from many people with a minimum cost. They are very useful in surveying user opinion, company procedures, individual practices and preferences, etc. Developing a useful questionnaire is not a simple process. There are experts available in questionnaire development and guidelines for developing and administering useful questionnaires (see Kidder, 1981).

The first rule of questionnaire design is that the questions should be simple and direct. The probability of confusing questions resulting in different people interpreting the questions or rating scales differently should be minimized. Confusing or ambiguous questions need to be eliminated. The best way to accomplish this is to administer the questionnaire to a small group of individuals who are part of the target population (e.g., pilots) and see how they interpret the questions. It is also very helpful to ask for their feedback on the format of the questionnaire, the clarity of the questions, etc.

While it is true that the best questions are simple and direct, care must be taken in the specific wording of the questions. A question with an obviously desirable answer will not yield informative results. For example, in a survey on cockpit and cabin crew coordination, Cardosi and Huntley (1988) wanted to assess crewmembers' knowledge of sterile cockpit procedures. The most direct question, "Do you know your airlines's procedure for sterile cockpit?" would probably have resulted in crewmembers answering in the affirmative, whether or not they were certain of the procedure. Instead, they asked, "What is your airline's procedure for sterile cockpit?" It was an interesting finding in itself that different crewmembers from the same airlines gave different answers.

Second, the questions need to be unbiased, both individually and as a set. Individual questions can be biased in terms of their wording. For example, asking "How much easier is it to use trackball X than trackball Y," presumes that trackball X is easier and respondents are unlikely to report that X is actually more difficult. An unbiased way to present the same question is "Compare the ease of using trackball X to using trackball Y." This question would be answered with a scale ranging from "X is more difficult" to "Y is more difficult" with a midpoint of "X and Y are the same."

Just as any individual question can be biased, a questionnaire may also be biased in its entirety. For example, if there are more questions about possible problems with a system than about its advantages, respondents may report feeling less favorably toward the system than if the questionnaire had more positive than negative questions.

Finally, the questionnaire should be administered as soon as possible after the experience or task that is under investigation. Because memory for detail can be very fleeting, it would not be advisable to show a pilot a new display, and then a week later administer the questionnaire. The sooner after exposure the questionnaire is administered, the more useful the results are likely to be. One exception to this rule is a questionnaire that is used to examine the effectiveness of a training program. That is, how much of the information that is presented in training is retained over a given period of time. For such a "test," a significant time interval (e.g., one month or longer) between exposure to the training and the questionnaire would be useful. A test with such a delay would be more effective than a test with no delay in predicting what information will be remembered and accessible for use when needed in actual operations.

Rating Scales

Rating scales are often very useful. Most scales offer five or seven choices. Fewer than five choices is confining; larger than seven, makes it difficult to define the differences between consecutive numbers on the scale.

Unless it is desirable to force questionnaire respondents to choose between two alternatives, rating scales should always have a mid-point. (This is one reason why an odd number of choices is recommended.) The scale should also have descriptive "anchors," that is, at least both ends and the middle values should have a word or phrase that identifies exactly what is meant by that number. This helps to minimize differences in people's own standards. For example, if the questions asks for a rating of the ease or difficulty of the use of a system Y as compared to system X, anchors should be given where the number '1' means much easier than X; the number '3' corresponds to 'no difference' and '5' means much more difficult than X. The results will be easier to interpret and, therefore, much more valuable, than those obtained by simply asking for a rating of ease or difficulty on a scale of one to five.

While user opinion is extremely valuable, there are many problems with making important design decisions by vote or consensus alone. We, as humans, are not very good at estimating our own response times, or predicting our own errors; nor do our initial preferences always match what will be most efficient in actual operations. Furthermore, there is also a tendency to prefer what is most familiar to us. Initial perceptions of new systems or subsystems may change with experience. For example, pilots who first used the B747-400 primary flight displays rated them as "very cluttered." With experience, however, these ratings change to "just right" (Boucek, personal communication). Also, the first line pilots to fly the B767 thought they preferred the electronic Horizontal Situation

Indicator (HSI), until they used the electronic map display (Boucek, personal communication).

It has also been the case that pilots have preferred one thing on the ground (e.g., a display with lots of high-tech options and information) and something else (usually a simpler, less cluttered, version) once they tried to use it in actual operations.

Even simple behaviors do not lend themselves to accurate judgments about our own actions. As part of an evaluation of a prototype navigation display, the Boeing flight deck integration team monitored pilot's eye movements as they used a prototype navigation display. The team also asked the pilots to report where they thought they were spending most of their time looking. There was no systematic relation between where the pilots thought they were looking the most and where the data actually showed that they were looking most (Fadden, personal communication).

Laboratory Experiments

It is difficult, if not impossible, to investigate issues by manipulating factors in actual operations. Such control is usually only available in a laboratory setting. The goal of an experiment is to manipulate the variables under investigation while keeping everything else constant. This careful manipulation of the key variables allows investigators to determine which of them has an effect.

One common type of a laboratory experiment is a part-task simulation. Part-task simulations are useful for studying simple questions, such as: "How long does it take to notice a particular change in the display?" or "Will the user immediately know what that symbols mean?" A part-task simulation is an ideal way to conduct an in-depth test of a new display. It allows attention to be focussed on the details of the display before it is tested operationally in a full-mission simulation. In addition to providing valuable results, a part-task simulation often points to specific areas that should be tested in a full-mission simulation.

The full-mission simulation is, of course, a very desirable type of test because it preserves the most realism, and thus, yields results that are easy to generalize to the real world. Full-mission simulation can give the same degree of control as a laboratory experiment, with the added benefits afforded by the realism.

The major drawback of full-mission simulation is that it is very expensive. The costs for computer time, simulator time, the salary for the pilots and/or controllers who participate, in addition to the other costs of research, can be prohibitive for all but the largest, and most well-funded, of projects. Also, there

are only a few places in the country that have the capability to conduct full mission simulation studies.

Another limitation of simulation studies that must be considered when interpreting the results is the priming effect. When pilots walk into a simulator knowing that they are going to participate in a test of Warning System X, they are expecting to see that system activated. They will see System X activated more times in one hour than they are likely to see in an entire day of line flying. This expectation leads to a priming effect which yields faster response times than can be expected when the activation of System X is not anticipated. For this reason, the response times obtained in simulations are faster than can be expected in the real-world and must be considered as examples of best-case performance. How much faster the response times will be in simulation than in actual operations is difficult to say as it depends on a variety of factors, particularly the specific task. In addition to response times being faster, they are also more homogeneous in simulation studies than would be expected in actual operations. This reduced variability can result in a higher likelihood of obtaining a statistically significant difference between two groups or conditions in a simulation study than in actual operations. However, since data obtained in actual operations are rarely obtainable, data from realistic simulation studies are a good alternative.

Experimental Validity and Reliability

The goal of any evaluation is to have reliable and valid results. Reliability refers to the repeatability of the results. If another investigator was to run the same test with the same equipment and same type of test participants, what are the chances that they would get the same results? In order to have repeatable results, the results obtained need to be due to the factors that were manipulated, and not to extraneous factors, chance, or anything peculiar to the testing situation or individuals tested.

In any experiment, it is necessary to carefully manipulate the factors that will be examined in the study and control all other variables (if only by keeping them constant). Careful controls help to ensure that the results of the study are, in fact, due to the factors examined and not to extraneous factors.

Validity refers to measuring what the test purports to measure. A classic example of this is the IQ test. Does it really measure one's ability to learn? Do the Standardized Aptitude Tests (SATs) actually measure one's ability to succeed in college? If the answer to this type of question is "no," then the test is not valid.

One way to help ensure that the results of the study are valid and reliable, is to employ careful controls of critical factors of interest and of extraneous factors (such as fatigued participants) that may influence the results of the study. This is easier said than done because it is often very difficult to even identify all of the factors that may contribute to your results. However, careful selection of test participants and testing conditions, in addition to a sound experimental design, will help to ensure valid and reliable results. A sound experimental design ensures that an adequate number of test participants ("subjects") are properly selected and tested (in an appropriate number and order of conditions) and that careful controls of the variables are included in the test.

Operationally Defined Variables

One fundamental component of an evaluation that often gets neglected is the idea that the test variables be operationally defined. This means that the factors under investigations must be defined in ways that can be measured. For example, a test to determine whether the use of the Traffic Alert and Collision Avoidance System (TCAS) increases Air Traffic Control (ATC) frequency congestion, would begin with an operational definition of frequency congestion. A suitable measure, in this case, would be the number of ATC calls generated by TCAS equipped aircraft (e.g., pilots contacting ATC to inform the controller of a maneuver or ask a question concerning a traffic alert) per unit time as compared to the number of traffic related calls generated by aircraft without TCAS under similar conditions.

Whether a test is designed to examine something simple (such as display clutter) or complex (such as situational awareness), all variables must be defined in terms of units that can be measured in the study.

Representative Subject Pool

Another necessary component of an evaluation is a representative subject pool. Since most research on basic perceptual and cognitive processes is conducted using college students as subjects, a question often arises as to whether or not we may generalize the results to specific populations, such as pilots or controllers.

One rule of thumb is that if the study purports to examine an aspect of behavior in which the target population would be expected to be different from college students in key ways, then the results will not be applicable. The differences between the target and test populations may be in terms of physical differences (such as age), or intellectual abilities (such as specific skills or knowledge). Whether or not these differences prevent a generalization of test

results depends upon the task. These differences can be quite subtle, but important.

For example, one approach to studying the similar call-sign problem might involve determining which numbers are most likely to be confused when presented auditorily. A sample research question would be "Is 225 more confusable with 252, or 235?". This is a relatively simple task and the results would comprise a confusability matrix. Because this is a simple auditory task, pilots would not be expected to perform much differently than college students (with the exception of the differences attributable to hearing loss due to age and exposure to noise). In this case, performance depends solely on the ability to hear the differences between numbers and results of experiments performed with college students as subjects are likely to be applicable to pilots.

Now consider a superficially similar, but technically very different, task. If the experimental task was to look at the effect of numerical grouping on memory for air traffic control messages, subjects might listen to messages with numerical information presented sequentially (e.g., "Descend and maintain one, zero thousand. Reduce speed to two two zero. Contact Boston Approach one one niner point six five"), and messages with numerical information presented in grouped form (e.g., "Descend and maintain ten thousand. Reduce speed to two twenty. Contact Boston Approach one nineteen *point sixty-five.") Since a pilot's* memory for that type of information is going to be very different from a college student's memory of that information, (mostly because it is meaningful to the pilot), results obtained by using college students would probably not be directly applicable to pilot populations.

One important aspect in which subjects should be representative of the target population is in terms of skill level. It is highly unlikely that a test pilot can successfully train himself to react or think like a line pilot. A below-average pilot (or an average pilot on a bad day) is likely to experience more difficulties with a new system than a skilled test pilot, or an Aircraft Evaluation Group (AEG) pilot. It is very difficult for a highly experienced operator to predict how people without prior knowledge or specific experiences will perform a certain task or what mistakes they are likely to make. Exceptional skill can enable an operator to compensate for design flaws - flaws which, because of the skill, may go unnoticed.

Controlling Subject Bias

While it is important that the people used as subjects are as similar as possible to the people to whom you want to generalize the results, it is also important that the subjects' biases don't affect the results of the test. If the participants

have their own ideas as to how the results should come out, it is possible for them to influence the results, either intentionally or unintentionally. It is not unusual for subjects to be able to discern the "desirable" test outcome and respond accordingly. To prevent this, investigators must take steps to control subject bias. For example, studies designed to test the efficacy of a new drug often employ a control group that receives a placebo (sugar pill). None of the subjects knows whether he or she is in the group receiving the new drug or in the group given the placebo. Some studies are conducted "double-blind" meaning that even the experimenters who deal with the subjects do not know who is receiving the placebo and who is receiving the drug.

In aviation applications, it is usually impossible to conduct a test (e.g., of new equipment) without the participants knowing the purpose of the test. Furthermore, this is often undesirable, since subjects' opinions (e.g., of the new display) can be a vital component of the data. One solution to the problem of controlling or balancing the effects of biases and expectations is the use of a control group. This group of subjects is tested under the same conditions (and presumably would have the same expectations) as the experimental group, but is not exposed to the tested variable.

For example, consider a test designed to examine the effectiveness of a new training program for wind shear (e.g., on the time required to maneuver based on a recognition of wind shear, number of simulated crashes, etc.). If the new training program is to be compared to an existing program, then the performance of pilots who were trained in each program could be compared. Pilots trained in the new program would be the experimental group and pilots trained in the existing program would constitute the control group. If the training program was a prototype and there was no such comparison to be made, then the performance of pilots trained with the new program (experimental group) could be compared to that of pilots who did not receive this training (control group). In this case, however, it would be important to control for test expectations. If, for example, the test wind shear scenarios were presented within days of the training, then the pilots would naturally expect wind shear to occur in the simulation sessions. This expectation would be expected to improve their performance over what it would be if wind shear was not anticipated. In this case, for the comparison between the two groups to be meaningful, pilots in both groups would need to be informed of the purpose of the test or be caught by surprise.

Another way to control subject bias is with careful subject selection. A good example of this is illustrated in a test conducted at the FAA Civil Aeromedical Institute to look at low-visibility minimums for passive auto-land systems (Huntley, unpublished study). The Air Transport Association (ATA) wanted

lower minimums than the Air Line Pilots' Association (ALPA) thought was safe. Clearly, both of these groups had a stake in the outcome of the test. When the simulation study was conducted, a portion of the subject pilots came from ALPA and an equal number of pilots came from ATA (Huntley, personal communication). While it is impossible to get rid of the biases that people bring to a test, it is usually possible to balance them out.

Representative Test Conditions

It is usually desirable for the test conditions to be as representative as possible to "real world" conditions. While the engineer looks at a system and asks, "Does it perform its intended function?", the human factors specialist wants to know if the pilots (or other operators) are able to use the system effectively under the conditions under which it will be used. Because of this, the key conditions included in the test must be as representative as possible to actual operating conditions so that the results of the test can be generalized to actual operations. Important conditions may include (but are not limited to): varied workload levels, weather conditions, ambient illumination levels (i.e., lighting conditions), ambient noise conditions, traffic levels, etc. For example, if a data input device is designed to be used in the cockpit, then it is important to ensure that it is easily used in a wide variety of lighting conditions and in turbulence (when it is difficult to keep a steady hand).

It is often important to include the "worst-case" scenario in addition to representative conditions in a test. Most human factors evaluations must include a worst case test condition, since it is the worst case (e.g., combination of failures) that often results in a dangerous outcome. For example, if it is important that a time-critical warning system be usable in all conditions, then the operator response time that is assumed by the software's algorithm needs to take this into account. In this case, in addition to measuring how long will it take the average person under average conditions to respond to the system, the longest possible response time, or response time at the 95th or 99th percentile, should also be measured. Such "worst case" response times should be obtained under "worst case" conditions.

Counter-balancing

One control that is not necessary in the engineering world but can be critical in the human factors world is counter-balancing. When measuring the noise level of two engines, it doesn't matter which one is tested first; the test of the first engine will not affect the outcome of the test of the second engine. When testing human performance, however, such order effects are common.

There are two possibilities of how human performance can change during the course of the test; it can get better or worse. Performance may improve because exposure to the first system gives subjects some information that helps them in using the second system. This is called positive transfer. For example, in a test of two data input devices, it would be reasonable to have pilots use each of them and measure the time required to perform specific tasks (response time) with each system. The number of errors made in the data input process (response accuracy) would also be measured. Performance with System A could be compared to performance with System B to determine which of the two systems is preferable. If the procedures for two systems are similar (e.g., in terms of keypad layout, the required order of the information input, etc.) but new to the pilot, then the practice acquired during test of System A might improve his or her performance with System B over what it would have been without the experience gained during the first test.

However, if the two systems are physically similar, but require different procedures to operate, then the experience acquired with the use (test) of System A would probably impair performance with System B. Performance with System B would have been better with no previous experience with System A. This phenomenon is referred to as negative transfer.

One way to avoid the possibility of positive or negative transfer influencing test results is to balance the order of conditions. For example, in a comparison of two navigation displays, a test could be conducted in which half of the pilots are tested with one display and half the pilots are tested using the other display. In this case, it is particularly important to ensure that there are no important differences in the two pilot populations (e.g., in terms of skill level). Alternatively, each pilot could be tested using both displays, with half of them using Display A first and half of them using Display B first. This is referred to as "counter-balancing."

There is another reason why performance may deteriorate over the course of a test. If the test is extremely long or the task is very tedious, performance may suffer due to a fatigue effect. When fatigue may be a factor in a test, careful controls (such as the use of an appropriate control group or balancing the order of conditions) must be considered. One study of the effects of fatigue on flight crews illustrates this point. Foushee, Lauber, Baetge, and Acomb (1986) investigated the effects of fatigue on flight crew errors. They had two groups of active line pilots fly a LOFT-type scenario in a full-mission simulation. Ten flightcrews flew the scenario within two to three hours after completing a three-day, high density, short-haul duty cycle. The other ten flightcrews flew the test scenario after a minimum of three days off. The results showed that while the "Post-Duty" crews were more fatigued than the "Pre-Duty" crews, their

performance was significantly better than that of the "Pre-Duty" crews. Of course, the better performance was not attributable to fatigue, but to a personal familiarity that developed over their duty cycle. The crews who had flown together on the duty cycle prior to the simulation got to know each other and knew what to expect from each other. This is often considered to be the birth of cockpit (or crew) resource management.

The first part of this study did not have the control group of pilots who flew together for the same amount of time right before the simulation, but weren't fatigued. As the second part of this study, a subsequent analysis of the data showed that the superior performance was, indeed, due to familiarity with the other crewmembers and not to fatigue.

How Should Test Results Be Analyzed?

Once the human factors test has been conducted, the next step is to analyze the results and present them in the simplest and most straightforward manner. The goals of data analysis are to describe the results and, where applicable, to determine whether there are important differences between groups or conditions of interest. Data analysis is used to summarize and communicate the meaning of a large set of numbers (such as response times or error rates) with the fewest possible numbers.

Descriptive Statistics

Measures of Central Tendency

Measures of central tendency seek to describe a set of data (e.g., a set of reaction times) with a single value. The most commonly cited measures of central tendency are the arithmetic mean, the median, and the mode.

The mean. The mean is computed as the sum of all the scores (e.g., response times or error rates) divided by the number of scores. For example, if response times (in msec.) for eight different pilots on a particular task were measured to be:

200, 225, 275, 300, 400, 400, 500, 1450

then the mean would be (200 + 225 + 275 + 300 + 400 + 400 + 500 + 1450)/8 or 469 msec. The mean is considered to be the fulcrum of a data set because the deviations in scores above it balances the deviation in scores below it. The sum of the deviations about the mean is always zero. Because of this, the mean is very sensitive to outlying scores, that is, scores that are very

different from the rest. A very high or very low score will tend to pull the mean in the direction of that score. In our example data set cited above, the mean of the first seven scores is 329 msec (compared to the mean of 469 with the score of 1450). While the mean is more frequently cited than the median or the mode, it is not always appropriate to cite it alone for this reason.

The median. The median is the score at which 50 percent of the scores fall above it and 50 percent of the scores are below it. With an odd number of scores, the median is the score in the middle when the scores are arranged from lowest to highest. With an even number of scores, the median is the average of the two middle scores. In the example array of data cited above, the median would be the average of 300 and 400 or 350 msec. One advantage of the median is that it is less sensitive to outlying data points. When there are a few scores that are very different from the rest, then the median should be considered as well as the mean.

The mode. The mode is the most frequently occurring score. In our example data set, the mode is 400, since it is the only score that occurs more than once. It is always possible, especially with very small data sets to have no mode. In very large data sets, it is possible to have multiple modes. While the mode is the most easily computed measure of central tendency, it is also less stable than the mean or median, and hence, usually not as useful.

Measures of Variability

A measure of central tendency, when presented in isolation, cannot fully describe the test results. In addition to the mean or median, we also need to know how close or disparate the scores were. In other words, how homogeneous were the scores as a group? For example, did half of the pilots take five seconds to perform the task and half of them require ten seconds or did they all take about 7.5 seconds? To answer this type of question, we need to compute a measure of variability, also known as a measure of dispersion. The most commonly used measure of dispersion is the standard deviation. The standard deviation takes into account the number of scores and how close the scores are to the mean.

The standard deviation (abbreviated as "s" or "s.d.") is the square root of the variance. The variance (s^2) equals the squared deviations of each score from the

mean divided by the total number of scores. One equation for computing the variance is as follows:

$$s^2 = \frac{\Sigma(X-\bar{X})^2}{n-1}$$

Where:

Σ is the summation sign
X represents each score
$\bar{X}$ equals the mean of the distribution, and
n equals the number of scores in the distribution.

To compute the standard deviation in this way, we subtract each score from the mean, square each difference, add the squares of the differences, divide this sum by the number of scores (or the number of scores minus one), and take the square root of the result. Relatively small standard deviation values are indicative of a homogeneous set of scores. If all of the scores are the same, for example, the standard deviation equals zero. In our sample set of data used to compute the mean, the standard deviation equals 383 msec.

Another use of the standard deviation is that it helps us to determine what scores, if any, we are justified in discarding from the data set. Studies in visual perception, for example, often use stimuli that are presented for very brief exposure durations (e.g., less than one-half of a second). In this case, a sneeze, lapse in attention, or other chance occurrence, could produce an extraordinarily long response time. This data point would not be representative of the person's performance, nor would it be useful to the experimenter. What objective criterion could be used to decide whether this data point should be included in the analysis?

In the behavioral sciences, it is considered acceptable to discard any score that is at least three standard deviations above or below the mean. In our sample set of data, if we discard the outlying score of 1450, the standard deviation becomes 100. Leaving this score out of the analysis would not be acceptable, however, using the convention of discarding scores three standard deviations above or below the mean. In this example, only scores above 1635 would be legitimately left out of the analysis. (In this case, it is impossible to have a score three standard deviations below the mean, because it would indicate a negative response time.)

Correlation

Correlation is a commonly used descriptive statistic that describes the relation between two variables. A correlation coefficient is reported as "r = x", where "x" equals some number between negative one and one. When two variables are unrelated (e.g., number of rainy days per month in Kansas and cost of airline fares), the correlation coefficient is near zero. A high positive "r" indicates that high values in one variable are associated with high values in the other variable. A high negative "r" indicates that high values in one variable are associated with low values in the other variable. A correlation of .7 or greater (or -.7 or less) is usually regarded as indicative of a strong relation between the two factors. An important note about correlation is that even a very high correlation (e.g., r =.90) does not imply causality or a cause-effect relationship. A correlation coefficient merely indicates the degree to which two factors varied together, perhaps, as a result of a third variable that remains to be identified.

Another way in which the correlation coefficient, is useful is that when squared, it indicates the percentage of the variance that is accounted for by the manipulated factors. For example, with a correlation coefficient of .7, the factors that were examined in the analysis account for only 49 percent of the variance (i.e., the variability in the data). The other 51 percent is due to chance or things that were not controlled.

Inferential Statistics

The statistics discussed above describe the test results and are, therefore, referred to as "descriptive statistics." Inferential statistics are used to determine whether two or more samples of data are significantly different, for example, if performance on System A is significantly better or worse than performance on System B.

The most commonly cited inferential statistics are the t-test and analysis of variance. Each method of analysis has an underlying set of assumptions. If these assumptions are seriously violated, or the analysis is inappropriate for the experimental design, then the conclusions based on the analysis are questionable.

Student's t Ratio

Student's t ratio (commonly referred to as a t-test) compares two different groups of scores and determines the likelihood that the differences found between them are due to chance. For example, t-tests would be appropriate when comparing the results of two groups of scores, whether it be the performance of the same group of pilots with System A and with System B, or

the performance of two groups of pilots - one using System A and the other using System B. When both sets of scores are taken from the same group of people, Student's t ratio for correlated samples is appropriate. When the scores of two different groups of people are examined, Student's t ratio for independent samples is appropriate. The formulas for computing a t-ratio (and all of the statistics discussed in this chapter) can be found in Experimental Statistics (Natrella, 1966) and in most statistics textbooks. Both types of t-tests look at the differences between the two groups of scores with reference to the variability found within the groups. They provide an indication as to whether or not the difference between the two groups of scores is statistically significant.

The results of a t-test are typically reported in the following format:

$t(df) = x, (p < p_o)$

Where:

"df" equals the number of degrees of freedom
"x" equals the computed t-value
"p_o" equals the probability value.

For example, $t(20) = 3.29, p < .01)$.

Degrees of freedom (df) refers to the number of values that are free to vary, once we have placed certain restrictions on the data. In the case of a t-test for correlated samples, the number of degrees of freedom equals the number of subjects minus one. For independent samples, df equals the number of subjects in one group added to the number of subjects in the other group minus two. In both cases, as the number of subjects increases (and, hence, the number of df increases), a lower t-value is required to achieve significance.

Statistical Significance

The p value relates to the probability that this specific result was achieved by chance. This is true not only for the t-values, but for all other statistics as well. A "$p < .01$" indicates that the probability that this result would be achieved by chance (and not due to the manipulated factors) is less than one in 100. When the results are significant at the .05 level, (i.e., $p \leq .05$), the chances of the results occurring by chance are 5 in 100, or less. Very often, the statistic is cited at the end of a statement of the results. For example, "The number of errors was significantly higher in the high workload condition than in the low workload condition ($t(15) = 2.25, p < .05$.)." It can also be used to show that there were no statistically significant differences between two conditions. For example, "The number of errors in the high workload conditions was

comparable to the number of errors in the moderate low condition ($t(15) = 0.92$, $p > .10$)." It cannot, however, be used to prove that there are no differences between the two groups, or that the two groups are the same.

For comparisons among more than two groups or more than two conditions in the same test, performing t-tests between all of the possible pairs would not be the best approach. A more appropriate test is Analysis of Variance (ANOVA). Analysis of Variance is similar to a t-test in that it examines the differences between groups with respect to the differences within groups. In fact, when there are only two groups, an analysis of variance yields the same probability value as the t-ratio.

Analysis of Variance

Analysis of variance (ANOVA) permits us to divide all of the potential information contained in the data into distinct, non-overlapping, components. Each of these portions reflects a certain part of the experiment, such as the effect of an individual variable (i.e., a main effect), the interaction of any of the variables, or the differences due solely to individual subjects. Each main effect and interaction is reported separately, in the following format:

$$F(df, df) = x, (p < p_o)$$

Where:

"df" equals the number of degrees of freedom
"x" equals the computed F statistic
"p_o" equals the probability value.

For example, $F(2,24) = 7.78$, $p < .01$). For an ANOVA, the two reported degrees of freedom are dependent upon the number of subjects and the number of conditions or levels of effects.

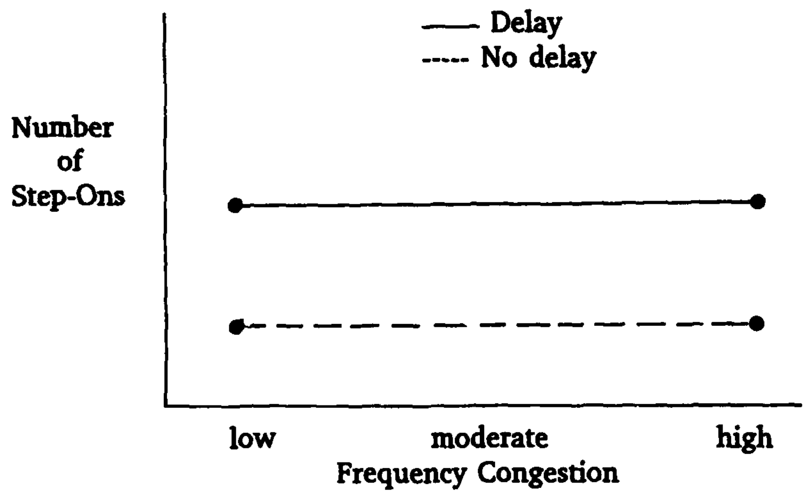
An Example

As a hypothetical example, consider a simulation study of the operational effects of transmitting pilot-to-controller and controller-to-pilot communications via satellites. (For an actual study that is very similar to the hypothetical one described here, see Nadler, et al., 1992.) This method of transmission would impose a delay (of approximately one-half second) between the time the controller keys the microphone and the time the pilot was able to hear the beginning of the transmission. Pilot transmissions to controllers would be similarly affected.

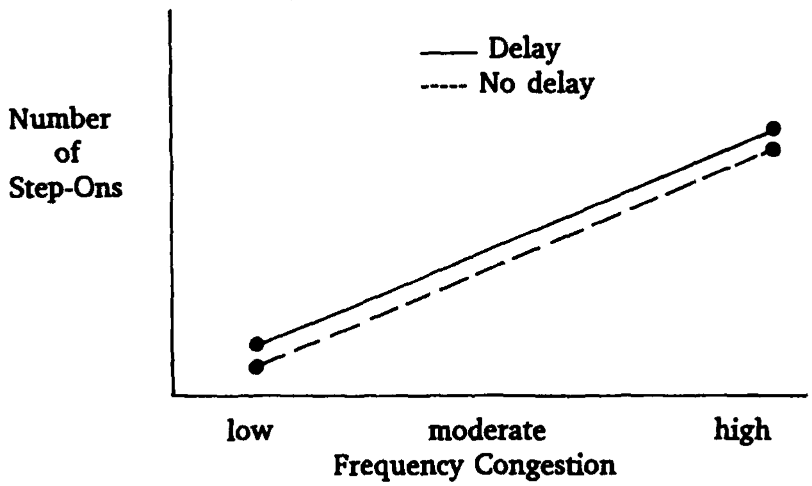
One effect that satellite transmission might be expected to have on operations is to increase the number of blocked transmissions ("step-ons"), since the delay makes it possible for both controllers and pilots to key their microphones without realizing that there is an incoming transmission. (The number of pilot-pilot step-ons would not change, as the pilots would still be able to hear the beginning of the other pilots' transmissions without a delay.) Without this delay induced by satellites, blocked transmissions are due solely to two or more people (controller and pilot or two pilots) attempting to transmit at the same time and to stuck mikes. Since the probability of two individuals trying to transmit simultaneously is logically a function of frequency congestion, the number of transmissions on the frequency would be an important experimental variable.

In this simulation study, two independent variables - the number of aircraft on the frequency and whether or not there is a communication delay - would be manipulated. Their effect on the number of step-ons (the dependent variable) would be measured. In this example, we have two levels of delay (500 msec. to simulate the satellite condition and no delay to simulate the present system). We are careful to ensure that the number of aircraft on the frequency generates different levels of frequency congestion. We categorize these levels of frequency congestion into "low," "moderate," and "high," based on data obtained from actual operations. Since we have two levels of delay and three levels of frequency congestion, this is referred to as a two-by-three experimental design. Furthermore, we have a completely balanced design. This means that we have an equal number of hours of voice recordings in each combination of delay-frequency congestion conditions. We are statistically confident that we have an adequate number of different controllers and number of hours of data. We are also careful to keep all other conditions constant (which is always easier said than done).

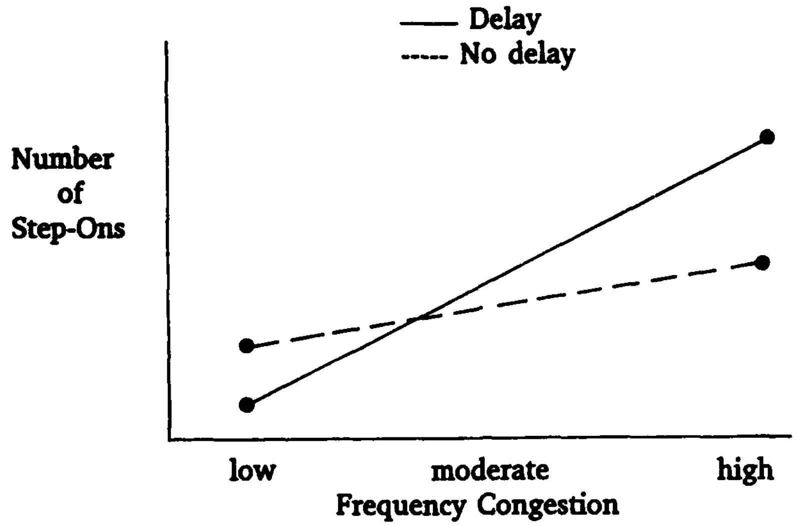
The three sources of variation in our ANOVA are the effects of delay, frequency congestion, and subjects (i.e., differences in the number of step-ons associated with different controllers). The results may show that the only significant effect is that of delay, meaning that the number of step-ons was significantly different for the two delay conditions. Graphically, this possibility might look like the following:



Another possible result is that the only significant effect was due to frequency congestion. This could mean that the number of step-ons increased with frequency congestion regardless of the delay condition. Graphically, this possibility might look like this:



In addition to one, both, or neither of the effects of delay and frequency congestion being significant, a significant interaction may occur. A significant interaction would occur if, for example, there was no difference between the delay conditions at the lowest level of frequency congestion, but there was a significant difference at the highest level of frequency congestion. Graphically, this possibility might look like the graph on the next page:



These are only examples of the type of results that may produce significant main effects or a significant interaction. There are many other possibilities. Of course, only a statistical analysis can determine whether differences portrayed on a graph are significant. Interpretation of test results is usually not simple, particularly with complex experimental designs. For this reason, human factors specialists with expertise in experimental design, but preferably statisticians, should be involved in the design of the research and the analysis of the results.

Regression Analysis

A special case of analysis of variance that is often used is regression analysis. Regression analysis takes the data and fits it to a mathematical function. The function may be a straight line, a parabola, or any other function. The analysis provides an indication of how well the data fits that particular function.

One of the advantages of regression analysis is that it is very forgiving of empty cells in an experimental design (i.e., conditions in the design that do not have as many data points as the other conditions). For example, if we wanted to test how many mistakes pilots were likely to make with a certain system, but were most interested in the number of errors to be expected under conditions of high workload, then we might run a test with the majority of responses being in high workload conditions. Perhaps some pilots would only be tested in the high workload condition. Because of this asymmetry of data points in the high and

moderate workload conditions, ANOVA would not be the most appropriate analysis; regression analysis, however, would still be appropriate.

Regression analysis also has some predictive value that analysis of variance does not. Regression analysis is often used to project from the data obtained in an experiment to situations that were not included in the test. In our hypothetical example, regression analysis would be appropriate if communication delays of 0 msec., 250 msec., and 500 msec. were tested and we wanted an estimate of the number of step-ons that could be expected at delays of 300 or 600 msec. When using regression analysis in this way, it is important to remember three points. First, the projection can only be as good as the fit of the data to the mathematical function. Second, all other things being equal, an estimate between two data points inspires more confidence than a projection beyond (above or below) the values included in the test. Third, confidence in the projection decreases as the distance between the hypothetical or projected point and the value that was included in the test increases.

Statistical vs. Operational Significance

A final note about data analysis concerns the differences betw en statistically significant and operationally significant results. Most statisticians only seriously consider results that are statistically significant at the .05 level or better. This enables the investigator to be reasonably certain that the findings were not due to chance. A statistically significant difference may, however, be very small as long as it is consistent. This may or may not be operationally useful. This difference between statistical significance and operational significance is often overlooked. A difference in response times of half of a second may be statistically significant, but may not be operationally important, depending upon the task.

On the other hand, when the experimental focus is actual operations, results that are not statistically significant at the .05 level may still be important. For example, if the focus of the experiment is serious operator errors that could significantly affect flight safety, then we may choose to conservatively consider results that are statistically significant only at the .1 level. The standard criteria for acceptance of statistical significance at the .05 level should not be used to ignore potentially interesting findings. It may also be the case that statistically significant results would be attainable with a more powerful test or change in research design (e.g., by utilizing better experimental controls or by increasing the number of subjects). The decision as to what level of significance is to be used should be dependent on the nature of the question that the test is designed to answer.

REFERENCES

Chapters 1-4

Albers, J. (1975) *Interaction of Color*. New Haven, CT: Yale University Press.

Alpern, M. (1971) Effector mechanisms in vision. In J.W. Kling & L.A. Riggs (Eds.) *Experimental Psychology*. New York: Holt, Rinehart & Winston, 367-394.

Appelle, S. (1972) Perception and discrimination as a function of stimulus orientation: The oblique effect in man and animals. *Psychological Bulletin*, 78, 266-278.

Bedford, R.E. & Wyszecki, G. (1958) Wave length discrimination for point sources. *Journal of the Optical Society of America*, 48, 129-135.

Békésy, G. & Rosenblith, W.A. (1951) The mechanical properties of the ear. In S.S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: John Wiley & Sons, Inc., 1075-1115.

Blakemore, C.B. & Sutton, P. (1969) Size adaptation: A new aftereffect. *Science*, 166, 245-247.

Boettner, E.A. (1967) *Spectral Transmission of the Eye* (data only). Report for contract AF41(609)-2966. University of Michigan. Prepared for USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas.

Bouman, M.A. & Walraven, P.L. (1957) Some color naming experiments for red and green monochromatic lights. *Journal of the Optical Society of America*, 47, 834-839.

Boynton, R.M (1978) Color in contour and object perception. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook of Perception*. New York: Academic Press.

Bowmaker, J.K., Dartnall, H.J.A. & Mollon, J.D. (1980) Microspectrophotometric demonstration of four classes of photoreceptors in an old world primate Macaca fascicularis. *Journal of Physiology*, 298, 131-143.

Brown, J.F. (1931) The thresholds for visual velocity. *Psychologische Forschung*, 14, 249-268.

Brown, J.L. (1965) Flicker and intermittent stimulation. In C.H. Graham (Ed.), ***Vision and Visual Perception***. New York: John Wiley & Sons, Inc., 251-320.

Campbell, F.W. & Robson, J.G. (1968) Application of Fourier analysis to the visibility of gratings. ***Journal of Physiology***, 197, 551-566.

Campbell, F.W. & Westheimer, G. (1960) Dynamics of accommodation responses of the human eye. ***Journal of Physiology***, 151, 285-295.

Campbell, F.W. & Wurtz, R.H. (1978) Saccadic omission: Why we do not see a grey-out during a saccadic eye movement. ***Vision Research***, 18, 1297-1303.

Carter, R.C. (1979) Visual search and color coding. ***Proceedings of the Human Factors Society 23rd Annual Meeting***, 369-373.

Chapanis, A. (1965) Color names for color space. ***American Scientist***, 53, 327-346.

Christ, R.E. (1975) Review and analysis of color coding research for visual displays. ***Human Factors***, 17, 542-570.

Cole, B.L. & Macdonald (1988) ***Defective Color Vision Can Impede Information Acquisition from Colour Coded Video Displays***. Department of Aviation Report No. 3, Melbourne, Australia.

Coren, S., Porac, C., & Ward, L.M. (1984) ***Sensation and Perception***. (2nd edit.) Orlando, FL: Harcourt Brace & Company.

Cornsweet, T.N. (1970) ***Visual Perception***. Orlando, FL: Harcourt Brace & Company.

Craik, F.I.M. & Lockhart, R.S. (1972) Levels of processing: A framework for memory research. ***Journal of Verbal Learning and Verbal Behaviour***, 11, 671-684.

Davis, R.G. & Silverman, S.R. (1960) ***Hearing and Deafness***. New York: Holt, Rinehart & Winston.

DeHaan, W.V. (1982) ***The Optometrist's and Ophthalmologist's Guide to Pilot's Vision***. Boulder, Co: American Trend Publishing Company.

deLange, H. (1958) Research into the nature of the human fovea-cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. ***Journal of the Optical Society of America***, 48, 777-784.

DeValois, R.L. & DeValois, K.K. (1988) ***Spatial Vision***. New York: Oxford University Press.

DeValois, R.L., Morgan, H. & Snodderly, D.M. (1974) Psychophysical studies of monkey vision--III. Spatial luminance contrast sensitivity tests of macaque and human observers. ***Vision Research***, 14, 75-81.

Ditchburn, R.W. (1955) Eye movements in relation to retinal action. ***Optica Acta***, 1, 171-176.

Dowling, J.E. & Boycott, B.B. (1966) Organization of the primate retina: Electron microscopy. ***Proceedings of the Royal Society of London***, Series B, 166, 80-111.

Durlach, I. & Colburn, H.S. (1978) Binaural phenomena. In E.C. Carterette & M.P. Friedman (Eds.), ***Handbook of Perception***. New York: Academic Press, 365-466.

Ericsson, K.A. & Faivre, I.A. (1988) What's exceptional about exceptional abilities? In L.K. Obler & D. Fein (Eds.), ***The Exceptional Brain: Neuropsychology of Talent and Special Abilities***, 436-473.

Evans, R.M. (1948) ***An Introduction to Color***. New York: John Wiley & Sons, Inc.

Fiorentini, A., Baumgartner, G., Magnussen, S., Schiller, P.H. & Thomas, J.P. (1990) The perception of brightness and darkness: Relations to neuronal receptive fields. In L. Spillmann & J.S. Werner (Eds.), ***Visual Perception: The Neurophysiological Foundations***. New York: Academic Press, 129-161.

Fletcher, H. (1953) ***Speech and Hearing in Communication***. (2nd edit.) New York: Van Nostrand Company.

Fletcher, H. & Munson, W.A., (1933) Loudness, its definition, measurement and calculation ***Journal of the Acoustical Society of America***, 5, 82-108.

Frisby, J.P. & Mayhew, J.E.W. (1976) Rivalrous texture stereograms. ***Nature***, 264, 53-56.

Fuld, K., Werner, J.S. & Wooten, B.R. (1983) The possible elemental nature of brown. ***Vision Research***, 23, 631-637.

Gibson, J.J. (1966) ***The Senses Considered as Perceptual Systems***. Boston: Houghton Mifflin.

Ginsburg, A.P., Evans, D.W., Sekuler, R. & Harp, S.A. (1982) Contrast sensitivity predicts pilots' performance in aircraft simulators. ***American Journal of Optometry & Physiological Optics***, 59, 105-108.

Goldstein, E.B. (1984) ***Sensation and Perception.*** Belmont, CA: Wadsworth Publishing Company, (2nd ed.).

Graham, C.H. (1965a) Some fundamental data. In C.H. Graham (Ed.), ***Vision and Visual Perception.*** New York: John Wiley & Sons, Inc., 68-80.

Graham, C.H. (1965b) Perception of movement. In C.H. Graham (Ed.), ***Vision and Visual Perception***, New York: John Wiley & Sons, Inc., 575-588.

Graham, C.H. & Brown, J.L. (1965) Color contrast and color appearances: Brightness constancy and color constancy. In C.H. Graham (Ed.), ***Vision and Visual Perception.*** New York: John Wiley & Sons, Inc., 452-478.

Ham, W.T., Mueller, H.A., Ruffolo, J.J., Guerry, D. & Guerry, R.K. (1982) Action spectrum for retinal injury from near-ultraviolet radiation in the aphakic monkey. ***American Journal of Ophthalmology***, 93, 299-306.

Hecht, S. (1934) Vision: II. The nature of the photoreceptor process. In C. Murchison (Ed.), ***A Handbook of General Experimental Psychology.*** Worcester, MA: Clark University Press, 704-828.

Hecht, S. & Smith, E.L. (1936) Intermittent stimulation by light: VI. Area and the relation between critical frequency and intensity. ***Journal of General Physiology***, 19, 979-989.

Hecht, S. & Verrijp, C.D. (1933) Intermittent stimulation by light. III. The relation between intensity and critical fusion frequency for different retinal locations. ***Journal of General Physiology***, 17, 251-265.

Helps, E.P.W. (1973) Physiological effects of aging. ***Proceedings of the Royal Society of Medicine***, 66, 815-818.

Hering, E. (1920; translation published 1964) ***Outlines of a Theory of the Light Sense.*** (trans. by L.M. Hurvich and D. Jameson; originally published in 1920). Cambridge, MA: Harvard University Press.

Higgins, K.E., Jaffe, M.J., Caruso, R.C. & deMonasterio, F.M. (1988) Spatial contrast sensitivity: Effects of age, test-retest, and psychophysical method. ***Journal of the Optical Society of America A***, 5, 2173-2180.

Hunt, G. (1982) Cathode ray tubes. In R.D. Bosman (Ed.) ***Modern Display Technologies and Applications.*** AGARD Advisory Report No. 169, 37-53.

Hurvich, L.M. (1981) ***Color Vision.*** Sunderland, MA: Sinauer Associates.

Iavecchia, J.H., Iavecchia, H.P. & Roscoe, S.N. (1988) Eye accommodation to head-up virtual images. ***Human Factors***, 30, 689-702.

Jameson, D. & Hurvich, L.M. (1972) Color adaptation: Sensitivity, contrast, after-images. In D. Jameson & L.M. Hurvich (Eds.), ***Handbook of Sensory Physiology***, Vol. VII/4 (pp. 568-581). Berlin: Springer-Verlag.

Judd, D.B. (1951) Basic correlates of the visual stimulus. In S.S. Stevens (Ed.), ***Handbook of Experimental Psychology.*** New York: John Wiley & Sons, Inc., 811-867

Julesz, B. (1971) ***Foundations of Cyclopean Perception.*** Chicago: University of Chicago Press.

Kaiser, P.K. (1968) Color names of very small fields varying in duration and luminance. ***Journal of the Optical Society of America***, 58, 849-852.

Kardon, D. (1989) Electroluminescent backlights for liquid crystal displays. ***Information Display***, 6, 17-21.

Kaufman, L. (1974) ***Sight and Mind: An Introduction to Visual Perception.*** New York: Oxford University Press.

Kaufman, L. & Rock, I. (1962) The moon illusion. ***Scientific American***, 207, 120-132.

Kelly, D.H. (1972) Adaptation effects on spatiotemporal sine wave thresholds. ***Vision Research***, 12, 89-101.

Kelly, D.H. (1974) Spatio-temporal frequency characteristics of color-vision mechanisms. ***Journal of the Optical Society of America***, 64, 983-990.

Kinney, J.A.S. (1983) Brightness of colored self-luminous displays. ***Color Research and Application***, 8, 82-89.

Klingberg, C.S., Elworth, C.S. & Filleau, C.R. (1970) Image quality and detection performance of military photointerpreters. ***Boeing Company Report*** D162-10323-1.

Krupka, D.C. & Fukui, H. (1973) The determination of relative critical flicker frequencies of raster-scanned CRT displays by analysis of phosphor persistence characteristics. *Proceedings of the Society for Information Display*, 14, 89-91.

Kuo, K.H. & Kalmanash, M.H. (1984) Automatic chrominance compensation for cockpit color displays. *SID 84 Digest*, 65-67.

Kurtenbach, W., Sternheim, C.E. & Spillmann, L. (1984) Change in hue of spectral colors by dilution with white light (Abney effect). *Journal of the Optical Society*, **A**, **1**, 365-372.

Laycock, J. (1982) *Recommended colours of use in airborne displays*. Technical Report 82110, Royal Aircraft Establishment, Farnborough, UK.

Leibowitz, H.W. (1983) A behavioral and perceptual analysis of grade crossing accidents. *Operation Lifesaver National Symposium 1982*. Chicago: National Safety Council.

Mach, E. (1865) über die Wirkung der räumlichen Verteilung des Lichtreizes auf die Netzhaut. *I. Sitzungsberichte der mathematisch-naturwissenschaftlichen Classe der kaiserlichen Akademie der Wissenschaften*, 52, 303-322.

Marc, R. E. & Sperling, H. G. (1977) Chromatic organization of primate cones. *Science*, 196, 454-456.

Matlin, M.W. (1983) *Sensation and Perception*. (2nd edit.) Boston: Allyn and Bacon.

McCollough, C. (1965) Color adaptation of edge-detectors in the human visual system. *Science*, 149, 1115-1116.

Meier, G. (1982) Electrochemical displays. In R.D. Bosman (Ed.) *Modern Display Technologies and* AGARD Advisory Report No. 169, 165-180.

Mikoshiba, S. (1988) Plasma displays. *Information Display*, 4, 14-18.

Murch, G.M. & Huber, J. (1982) Colour--the logical next step. *New Electronics*, 15, 31-32.

Newman, C.V., Whinham, E.A. & MacRae, A.W. (1973) The influence of texture on judgment of slant and relative distance in a picture with suggested depth. *Perception & Psychophysics*, 14, 280-284.

Nickerson, D. & Newall, S.M. (1943) A psychological color solid. ***Journal of the Optical Society of America***, 33, 419-422.

Osterberg, G. (1935) Topography of the layer of rods and cones in the human retina. ***Acta Ophthalmologica*, Supplement 6.**

Owsley, C., Sekuler, R. & Siemsen, D. (1983) Contrast sensitivity throughout adulthood. ***Vision Research***, 23, 689-699.

Pearlman, A.L., Birch, J. & Meadows, J.C. (1979) Cerebral color blindness: An acquired defect in hue discrimination. ***Annals of Neurology***, 5, 253-261.

Pettigrew, J.D. (1972) The neurophysiology of binocular vision. ***Scientific American***, 227, 84-96.

Pitts, D.G. (1982) The effects of aging on selected visual functions: Dark adaptation, visual acuity, stereopsis and brightness contrast. In R. Sekuler, D.W. Kline and K.Dismukes (Eds.), ***Aging and Human Visual Functions***. New York: Liss.

Pokorny, J, Smith, V.C., Verriest, G. & Pinckers, A.J.L.G. (1979) ***Congenital and Acquired Color Vision Defects***. New York: Grune & Stratton.

Polyak, S. (1941) ***The Retina***. Chicago: University of Chicago Press.

Purdy, D.M. (1931) Spectral hue as a function of intensity. ***American Journal of Psychology***, 43, 541-559.

Regan, D., Beverly, K. & Cynader, M. (1979) The visual perception of motion in depth. ***Scientific American***, 241, 136-151.

Richards, W. (1970) Stereopsis and stereoblindness. ***Experimental Brain Research***, 10, 380-388.

Riggs, L.A. Visual acuity. In C.H. Graham (Ed.), ***Vision and Visual Perception***. New York: John Wiley & Sons, Inc., 321-349.

Riggs, L.A., Ratliff, F., Cornsweet, J.C. & Cornsweet, T.N. The disappearance of steadily fixated visual test objects. ***Journal of the Optical Society of America*, 43,** 495-501.

Riggs, L.A., Volkmann, F.C. & Moore, R.K. (1981) Suppression of the blackout due to blinks. ***Vision Research***, 21, 1075-1079.

Rovamo, J., Virsu, V. & Näsänen, R. (1978) Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. ***Nature***, 271, 54-56.

Schlam, E. (1988) Thin-film electroluminescent displays. ***Information Display***, 4, 10-13.

Sekuler, R. (1974) Spatial vision. ***Annual Review of Psychology***, 25, 195-232.

Sekuler, R. & Blake, R. (1985) ***Perception***. New York: McGraw Hill, Inc.

Sherr, S. (1979) ***Electronic Displays***. New York: John Wiley & Sons, Inc.

Snyder, H.L. (1980) ***Human Visual Performance and Flat Panel Display Image Quality***. Technical Report HFL-80-1/ONR-80-1.

Snyder, H.L. (1988) Image Quality. In M. Helander (Ed.), ***Handbook of Human-Computer Interaction***. Amsterdam: Elsevier Science Publishers.

Society of Automotive Engineers (1988) ***Human Engineering Considerations in the Application of Color to Electronic Aircraft Displays***. ARP4032. Warrendale, PA: SAE.

Stokes, A.F. & Wickens, C.D. (1988) Aviation displays. In E.L. Wiener & D.C. Nagel (Eds.), ***Human Factors in Aviation***. New York: Academic Press.

Taylor, C.A. (1965) ***The Physics of Musical Sounds***. New York: Elsevier.

Teichner, W.H. (1979) Color and visual information coding. ***Proceedings of the Society for Information Display***, 20, 3-9.

Tolin, P. (1987) Maintaining the three-dimensional illusion. ***Information Display***, 3, 10-12.

Turnage, R.E. (1966) The perception of flicker in cathode ray tube displays. ***Information Display***, 3, 38.

Uchikawa, H., Kaiser, P. K. & Uchikawa, K. (1982) Color-discrimination perimetry. ***Color Research and Application***, 7, 264-272.

Verriest, G. (1963) Further studies on acquired deficiency of color discrimination. ***Journal of the Optical Society of America***, 53, 185-195.

Vision Committee, National Research Council (1981) ***Procedures for Testing Color Vision.*** Report of Working Group 41. Washington, D.C.: National Academy Press.

Viveash, J.P. & Laycock, J. (1983) Computation of the resultant chromaticity coordinates and luminance of combined and filtered sources in display design. ***Displays***, 4, 17-23.

Volbrecht, V.J., Aposhyan, H.M. & Werner, J.S. (1988) Perception of electronic display colours as a function of retinal illuminance. ***Displays***, 9, 56-64.

Volkmann, F.C. (1962) Vision during voluntary saccadic eye movements. ***Journal of the Optical Society of America***, 52, 571-578.

Vos, J. J. & Walraven, P. L. (1971) On the derivation of the foveal cone primaries. ***Vision Research***, 11, 799-818.

Wald, G. (1945) Human vision and the spectrum. ***Science***, 101, 653-658.

Walraven, J. (1985) The colours are not on the display: a survey of non-veridical perceptions that may turn up on a colour display. ***Displays***, *6, 35-42.*

Walraven, J., Enroth-Cugell, C., Hood, D.C., MacLeod, D.I.A. & Schnapf, J.L. (1990) The control of visual sensitivity: Receptoral and postreceptoral processes. In L. Spillmann & J.S. Werner (Eds.), ***Visual Perception: The Neurophysiological Foundations***, New York: Academic Press, 53-101.

Walraven, P.L. (1962) ***On the Mechanisms of Colour Vision.*** Ph.D. Thesis, University of Utrecht.

Ward, W.D. & Glorig, A. (1961) A case of fire-cracker induced hearing loss. **Laryngoscope**, 71, 1590-1596.

Weale, R.A. (1982) ***A Biography of the Eye.*** London: H.K. Lewis & Co.

Weale, R.A. (1988) Age and the transmittance of the human crystalline lens. ***Journal of Physiology***, 395, 577-587.

Welch, R.B. & Warren, D.H. (1980) Immediate perceptual response to intersensory discrepancy. ***Psychological Bulletin***, 88, 638-667.

Werner J.S. (1982) Development of scotopic sensitivity and the absorption spectrum of the human ocular media. ***Journal of the Optical Society of America***, 72, 247-258.

Werner, J.S., Cicerone, C.M., Kliegl, R. & DellaRosa, D. (1984) Spectral efficiency of blackness induction. ***Journal of the Optical Society of America***, 1, 981-986.

Werner, J.S., Peterzell, D. & Scheetz, A.J. (1990) Light, vision, and aging -- A brief review. ***Optometry and Vision Science***, 67 214-229.

Werner, J.S. & Schlesinger, K. (1991) ***Psychology: Science of Mind, Brain, and Behavior***. New York: McGraw-Hill.

Werner, J.S. & Steele, V.G. (1988) Sensitivity of human foveal color mechanisms throughout the life span. ***Journal of the Optical Society of America***, A, 5, 2122-2130.

Werner, J.S. & Walraven, J. (1982) Effect of chromatic adaptation on the achromatic locus: The role of contrast, luminance and background color. ***Vision Research***, 22, 929-943.

Werner, J.S. & Wooten, B.R. (1979) Opponent-chromatic mechanisms: Relation to photopigments and hue naming. ***Journal of the Optical Society of America***, 69, 422-434.

Wertheimer, M. (1912) Experimentelle Studien über das Sehen von Bewegung. ***Zeitschrift für Psychologie***, 61, 161-265.

Westheimer, G. (1964) Pupil size and visual resolution. ***Vision Research***, 4, 39-45.

Wheatstone, C. (1838) Some remarkable phenomena of binocular vision. (Reprinted in 1964) In W.N. Dember (Ed.), ***Visual Perception: The Nineteenth Century***, New York: John Wiley & Sons, Inc., 114-129.

Williams, R.D. & Garcia, F. (1989) Volume visualization displays. ***Information Display***, 5, 8-10.

Wilson, H.R., Levi, D., Maffei, L., Rovamo, J. & DeValois, R. (1990) The perception of form: Retina to striate cortex. In L. Spillmann & J.S. Werner (Eds.), ***Visual Perception: The Neurophysiological Foundations***. New York: Academic Press, 231-272

Wooten, B.R. (1984) The effects of successive chromatic contrast on spectral hue. In L. Spillmann & B.R. Wooten (Eds.), ***Sensory Experience, Adaptation, and Perception***. (pp. 471-494) Hillsdale, NJ: Lawrence Erlbaum Associates.

Wright, W.D. (1946) ***Researches on Normal and Defective Colour Vision***. London: Henry Kimpton.

Wright, W.D. & Pitt, F.H.G. (1934) Hue discrimination in normal colour-vision. ***Proceedings of the Physical Society***, 46, 459-473.

Wurtz, J.E. (1989) The not-so-amazing survival of the CRT. ***Information Display***, 5, 5-18.

Wyszecki, G. & Stiles, W.S. (1982) ***Color Science: Concepts and Methods, Quantitative Data and Formulae***. (2nd ed.) New York: John Wiley & Sons, Inc.

Yarbus, A.L. (1967) ***Eye Movements and Vision***. New York: Plenum Publishing Corp.

Young, R.W. (1988) Solar radiation and age-related macular degeneration. ***Survey of Ophthalmology***, 32, 252-269.

Zelman, S. (1973) Correlation of smoking history with hearing loss. ***Journal of the American Medical Association***, 223, 920.

Zihl, J., von Cramon, D. & Mai, N. (1983) Selective disturbance of movement vision after bilateral brain damage. ***Brain***, 106, 313-340.

Zrenner, E., Abramov, I. Akita, M., Cowey, A., Livingstone, M. & Valberg, A. (1990) Color perception: Retinal, geniculate, and cortical mechanisms. In L. Spillmann & J.S. Werner (Eds.), ***Visual Perception: The Neurophysiological Foundations***. New York: Academic Press, 163-203.

Zwicker, E. (1958) Uber psychologische und methodosche grundlagen der Lautheit. ***Acustica***, 8, 237-258.

Chapter 5

Bransford, J. D., Barclay, J. R. & Franks, J.J. (1972) Sentence memory: A constructive versus interpretive approach. ***Cognitive Psychology***, 3, 193-209.

Craik, F. & Lockhart R. (1972) Level of Processing: A Framework for Memory Research. ***Journal of Verbal Learning and Verbal Behavior***, 11, 671-684.

James, W. (1890) *The Principles of Psychology*, New York: Holt.

Johnston, W. A., & Heinz, S. P. (1978) Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General*, 107, 420-435.

Kahneman, D. *Attention and Effort*. (1973) Englewood Cliffs, NJ: Prentice Hall.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., & Fujimura, O. (1975) An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18, 331-340.

Nadler, E., Mengert, P., Sussman, E.D., Grossberg, M., Salomon, A., & Walker, K., (unpublished manuscript). Effect of Binaural Delays in the Communication Channel Linking Radar and Data Controller.

Neisser, U. (1967) *Cognitive Psychology*, New York: Appleton-Century-Crofts.

Neisser, U. & Becklen, R. (1975) Selective Looking: Attending to visually specified events. *Cognitive Psychology*, 7, 480-494.

Palmer, S. E. (1975) The effects of contextual scenes on the identification of objects. *Memory and Cognition*, 3, 519-526.

Peterson, G. E. & Barney, H.L. (1952) Control methods used in the study of vowels. *Journal of the Acoustical Society of America*, 24, 175-184.

Reicher, G. M. (1969) Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275-280.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987) Focused attention and frequency selectivity. *Perception & Psychophysics*, 42, 215-221.

Sperling, G. S. (1960) The information available in brief visual presentations. *Psychological Monographs*, 74, 1-29.

Tsal, Y. (1983) Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 523-530.

Warren, R.M. (1970) Perceptual Restoration of Missing Speech Sounds. *Science, 167, 392-393.*

Warren, R.M. & Obusek, C.J. (1971) Speech Perception and Phonemic Restorations. *Perception and Psychophysics*, 9 (38), 358-362.

Weisstein, N. & Harris, C. S. (1974) Visual detection of line segments: An object-superiority effect. ***Science***, 186, 752-755.

Chapters 6-8

Aldrich T.B., Szabo, S.M., & Bierbaum, C.R. (1989) The development and application of models to predict operator workload during system design. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton, & L. Van Breda (Eds.), ***Applications of Human Performance to System Design***, New York: Plenum Publishing Corp., 65-80.

Allnut, M.F. (1987). Human factors in accidents. ***British Journal of Anaesthesia***, 59, 856-864.

Andre, A.D. & Wickens, C.D. (1990). Display control compatability in the cockpit: guidelines for display layout analysis. ***Tech Report, ARL-90-12/NASA/A3I-90-1.*** Savoy, Illinois: University of Illinois Aviation Research Lab.

Braune, R.J. (1989) The common/same type rating: human factors and other issues. ***Paper No. 892229***, Warrendale, PA: SAE.

Cooper, G.E. & Harper, R.P., Jr. (1969) The use of pilot rating in the evaluation of aircraft handling qualities. ***NATO: AGARD-AG-567.*** Paris, France (DTIC).

Czeisler, C.A., Weitzman, E.D., Moore-Ede, M.C., Zimmerman, J.C., & Knauer, R.S. (1980) Human sleep: its duration and organization depend on its circadian phase. ***Science***, 210, 1264-1267.

Degani, A. & Weiner, E.L. (1990) Human factors of flight deck checklists: the normal checklist. ***NASA Contractor Report 177549,*** NASA Ames Research Center, Moffet Field, CA.

Desmond, J. (1986) Improvements in aircraft safety and operational dependability from a projected flight path guidance display. ***Paper No. 861732,*** Warrendale, PA: SAE.

Farmer, E.W. & Green, R.G. (1985) The sleep-deprived pilot: performance and EEG response. ***16th Conference for Western European Association for Aviation Psychology,*** Helsinki, 1985, 24-285.

Fischer, E., Haines, R. & Price, T. (1980, December) Cognitive Issues in Head-Up Displays, ***NASA Technical Paper 1171,*** NASA, Washington, DC.

Gopher, D. (1991) The skill of attention control: acquisition and execution of attention strategies. In D. Meyer & S. Kornblum (Eds.), ***Attention and Performance IVX***, Hillsdale, NJ: Erlbaum Associates.

Graeber, R.C. (1988) Aircrew fatigue and circadian rhythmicity. In E. Weiner & D. Nagel (Eds.), ***Human Factors in Aviation***, New York: Academic Press, 305-344.

Graeber, R.C. (1988, 1989) Jet lag and sleep disruption. In M.H. Kryger, T. Roth, & W.C. Dement (Eds.), ***Principles and Practice of Sleep Medicine***, Philadelphia, PA: W.B. Saunders Company, 324-331.

Griffin & Rockwell (1989) The confirmation bias is supported by ***expectancy***. What you expect to see helps you confirm what you believe your state is. A major concern in private pilot aviation is continued flight into deteriorating weather. This has been documented by research at Ohio State University (Griffin & Rockwell, 1989).

Groce, J.L. & Boucek, G.P., Jr. (1987). Air transport crew tasking in an ATC data link environment. ***Paper No. 871764.*** Warrendale, PA: SAE.

Hart, S.G. (1989) Crew workload management strategies: a critical factor in system performance. ***Fifth Annual Symposium on Aviation Psychology***, Columbus, OH.

Hartzell, E.J., Dunbar, S., Beveridge, R. & Cortilla, R. (1982) Helicopter pilot response latency as a function of the spatial arrangement of instruments and controls. ***Proceedings of the 28th Annual Conference on Manual Control***, Wright-Patterson AFB, Dayton, OH.

Hawkins, F.H. (1987) ***Human Factors in Flight***, Brookfield, VT: Gower Technical Press.

Hyman, R. (1953) Stimulus information as a determinant of reaction time. ***Journal of Experimental Psychology***, 45, 423-432.

Jensen, R.J. & Benel, R. (1977) Judgment evaluation and instruction in civil pilot training. ***Final Report FAA-RD-78-24***, Springfield, VA, National Technical Information Service.

Johnson, S.L. & Roscoe, S.N. (1972) What moves, the airplane or the world? ***Human Factors***, 14, 107-129.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.), (1982) ***Judgment Under Uncertainty: Heuristics and Biases***, New York: Cambridge University Press.

Klein, G.A. (1989a) Do decision biases explain too much? ***Human Factors Society Bulletin***, 32, 1-3.

Klein, G.A. (1989b) Recognition-primed decisions. In W. Rouse (Ed.) ***Advances in Man-Machine Systems Research***, Vol. 5, Greenwich, CT: JAI Press, 47-92.

Klein, K.G., Wegmann, H.M. & Hunt, B.I. (1972) Desynchronization of body temperature and performance circadian rhythms as a result of outgoing and homegoing transmeridian flights. ***Aerospace Medicine***, 43, 119-132.

Nagel, D.C. (1988) Human error in aviation operations. In E. Weiner & D. Nagel (Eds.), ***Human Factors in Aviation***, New York: Academic Press, 263-303.

Newman, R.L. (1987) ***Improvement of head-up display standards:*** Volumes I, II, III. (AFWAL-TR-3055). Wright-Patterson Air Force Base, OH: Flight Dynamics Laboratory.

Norman, D. (1988) ***The Psychology of Everyday Things***, New York: Harper-Collins.

North, R.A., & Riley, V.A. (1989) A predictive model of operator workload. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton, & L. Van Breda (Eds.), ***Applications of Human Performance to System Design***, New York: Plenum Publishing Corp., 81-99.

Parks, D.L. & Boucek, G.P., Jr. (1989) Workload prediction, diagnosis and continuing challenges. In G.R. McMillan et al (Eds.), ***Applications of Human Performance Models to System Design***, New York: Plenum Publishing Corp., 47-63.

Reason, J. (1990) ***Human Error***, New York: Cambridge University Press.

Richardson et al. (1982) Circadian Variation in Sleep Tendency in Elderly and Young Subjects. ***Sleep***, 5 (suppl.2), A.P.S.S., 82.

Roscoe, A.H. (1987). ***The practical assessment of pilot workload.*** NATO: AGARD-AG-282. Loughton, U.K.: Specialized Printing Services Ltd.

Roscoe, S.N. (1968) Airborne displays for flight and navigation. ***Human Factors***, 10, 321-322.

Steenblik, J.W. (1989, December) Alaska Airlines' HGS. ***Air Line Pilot***, 10-14.

Vicente, K.J., Thornton, D.C., & Moray, N. (1987) Spectral analysis of sinus arrhythmia: a measure of mental effort. ***Human Factors***, 29(2), 171-182.

Wegman, H.M., Gundel, A., Naumann, M., Samel, A., Schwartz, E. & Vejvoda, M. (1986). Sleep, sleepiness, and circadian rhythmicity in aircrews operating on transatlantic routes. ***Aviation, Space, and Environmental Medicine***, 57, (12, suppl.), B53-B64.

Weiner, E.L. (1988) Cockpit automation. In E.L. Weiner & D.C. Nagel (Eds.), ***Human Factors in Aviation***, New York: Academic Press, 263-303.

Weinstein, L.F. (1990) The reduction of central-visual overload in the cockpit. ***Proceedings of the 12th Symposium on Psychology in the Department of Defense***, U.S. Air Force Academy, Colorado Springs.

Weintraub, D.J. & Ensing, M. (1992) ***Human Factors Issues in Head-Up Display Design: Book of HUD - CSERIAC.*** Wright-Patterson Air Force Base, OH: Crew System Ergonomics Information Analysis Center (CSERIAC).

Weintraub, D.J., Haines, R.F. & Randall, R.J. (1984) The utility of head-up displays: eye focus vs. decision times. ***Proceedings of the 28th Annual Meeting of the Human Factors Society***, 1, 529-533.

Wickens, C.D. (1984). ***Engineering Psychology and Human Performance*** (1st edit.), New York, Harper/Collins.

Wickens, C.D. (1992). ***Engineering Psychology and Human Performance*** (2nd edit.), New York, Harper/Collins.

Wickens, C.D & Flach, J. (1988) Human Information Processing, in E. Weiner & D. Nagel (Eds.), ***Human Factors in Aviation***, 111-155, New York: Academic Press.

Wickens, C.D., Martin-Emerson, R., & Larish, I. (1993) Attentional tunneling and the head-up display. ***Proceedings from the 7th International Symposium on Aviation Psychology.*** Columbus, Ohio: Ohio State University, Department of Aviation.

Wickens, C.D., Stokes, A.F., Barnett, B. & Hyman, F. (1988). Stress and pilot judgment: an empirical study using MIDIS, a microcomputer-based simulation. ***Proceedings of the 32nd Meeting of the Human Factors Society***, Human Factors Society, Santa Monica, CA.

Wilson, G.F., Skelly, J., & Purvis, G. (1989). Reactions to emergency situations in actual and simulated flight. In ***Human Behavior in High Stress Situations in Aerospace Operations.*** NATO: AGARD-CP-458. Loughton. U.K.: Specialized Printing Services Ltd.

Wright, P. (1977) Presenting technical information: a survey of research findings. ***Instructional Science***, 6, 93-134.

Chapters 9-11

Bainbridge, L. (1987) Ironies of automation. In ***New Technology and Human Error***. Ed. M.J. Rasmussen, K. Duncan, and J. Leplat. West Sussex, U.K.: John Wiley & Sons Ltd.

Barnett, A. & Higgins. M.K. (1989) Airline safety: the last decade. ***Journal of Management Science***, 35, January 1989, 1-21.

Broadbent, D.E. (1971) ***Decision and Stress***. London: Academic Press.

Card, J. K. (1987. Conversation with Richard Gabriel, Palo Alto, California.

Corwin, W.H., Sandry-Garza, D.L., Biferno, M.H., Boucek, G.P., Logan, A.L., Jonsson, J.E., & Metalis, S.A. (1989) ***Assessment of crew workload measurement methods, techniques, and procedures*** (WRDC-7R-89-7006). Wright-Patterson Air Force Base, Wright Research and Development Center, Dayton, Ohio.

Crossman, E.R., Cooke, J.E., & Beishon, R.J. (1974) Visual attention and sampling of displayed information in process control. In ***The Human Operator in Process Control.*** Ed. M. E. Edwards and J. Lees, 25-50.

Curry, R.E. (1985) The introduction of new cockpit technology: a human factors study. ***NASA Technical Memorandum 86659.*** Moffett Field, California: NASA Ames Research Center.

Department of Defense. (1985) ***Human engineering guidelines for management information systems*** (DOD-HBK-761). Wright-Patterson Air Force Base, Air Force Systems Command, Dayton, Ohio.

Fadden, D.M., & Weener, E.F. (1984) Selecting effective automation. Paper presented at ALPA Air Safety Workshop Washington, D.C.

Gabriel, R.F. (1987) Internal memorandum. Society of Automotive Engineers (SAE) Committee on Human Behavioral Technology.

Hilgard, E.R., & Atkinson, R.C. (1967) *Introduction to Psychology*. New York: Harcourt, Brace and World.

Meister, D. (1987) A cognitive theory of design and requirements for a behavioral design aid in system design. In *Behavioral Perspectives for Designers, Tools and Organizations*. Ed. W.B. Rouse and K.R. Boff. New York: North Holland Publications.

Nagel, D.C. (1987) Aviation safety: needs for human factors research. *Presentation to Air Transport Association (ATA) Airlines Operations Forum*, 19-21 October, Annapolis, Maryland.

Norman, S.D. & Orlady, H.W. (Eds.) (1988) *Flight Deck Automation: Promises and Realities Final Report of a NASA/FAA/Industry Workshop*. Moffett Field, California: NASA Ames Research Center.

Rogers, Locan, & Boley (1989) *Classification and reduction of pilot error*. NASA, CR-181867, DOT/FAA/DS-89/24. Hampton, VA: NASA.

Roth, E.M., Bennett, K.B., & Woods, D.D. (1987) Human interaction with an intelligent machine. *International Journal of Man-Machine Studies* (27).

Rouse, W.B. (1977) Human-Computer interaction in multitask situations. *IEEE Transactions on Systems, Man and Cybernetics*, 384-392.

Rouse W.B., & Boff, K.R. (1987) System design. In *Behavioral Perspectives on Designers, Tools, and Organizations*. New York: North Holland Publications.

Sheridan, T.B. (1980) Computer control and human alienation. *Technology Review*, 61-73.

Sinaiko, H.W. (1972) Human intervention and full automation in control-systems. *Applied Ergonomics*, 3-7.

Society of Automotive Engineers (1990) *Human interface design methodology for integrated display symbology*. ARP 4155. Warrendale, PA: SAE.

Webster's New World Dictionary, (1970) Guralnik, D.B. (Ed.) New York: World Publishing Co.

Wiener, E.L. (1985) Human factors of cockpit automation: A field study of flight crew transition *NASA Contractor Report 177333*. Moffett Field, California: NASA Ames Research Center.

Woods, D.D. (1987) ***Technology alone is not enough: reducing the potential for disaster in risky technologies.*** Pittsburgh, Pennsylvania: Westinghouse Research and Development Center.

Chapter 12

Berson, B. L., Po-Chedley, D. A., Boucek, G. P., Hanson, D. C., Leffler, M. F., & Wasson, R. L. (1981) ***Aircraft Alerting Systems Standardization Study Volume II: Aircraft Alerting System Design Guidelines***, Report No. DOT/FAA/RD-81/II.

Boff, K. R., & Lincoln, J. E., (Eds.) (1988) ***Engineering Data Compendium: Human Perception and Performance,*** Volume III, Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, 1862.

Boucek, G. P., Erickson, J. B., Berson, B. L., Hanson, D. C., Leffler, M. F., & Po-Chedley, D. A., (1980) ***Aircraft Alerting Systems Standardization Study Phase I.*** Final Report. Report No. FAA/RD-80-68.

Boucek, G. P., Pfaff, T. A., White, R. W., & Smith, W. D. (1985) ***Traffic Alert and Collision Avoidance System - Operational Simulation***. Report No. DOT/FAA/PM-85/10.

Boucek, G. P., Po-Chedley, D. A., Berson, B. L., Hanson, D. C., Leffler, M. F., & White, R. W. (1981) ***Aircraft Alerting Systems Standardization Study, Volume I: Candidate System Validation and Time-Critical Display Evaluation***. Report No. FAA-RD-81-381.

Boucek, G. P., White, R. W., Smith, W. D., & Kraus, J. M. (1982) ***Traffic Alert and Collision Avoidance System - Developmental Simulation***, Report No. DOT/FAA/RD-82/49.

Cardosi. K. & Boole, P. (1991) ***Analysis of Pilot Response Time to Time-Critical Air Traffic Control Calls.*** Report No. DOT/FAA/RD-91/20.

Cardosi, K. & Huntley, M. S. (1988) ***Cockpit and Cabin Crew Coordination.*** Report No. DOT/FAA/FS-88/1.

Foushee, C., Lauber, J., Baetge, M., & Acomb, D. (1986) ***Crew Factors in Flight Operations III; The operational significance off exposure to short-haul air transport operations.*** NASA Technical Memorandum 88322 Moffett field, CA.

Hopkin, D. (1980) The measurement of the air traffic controller. ***Human Factors,*** 22(5).

Kidder, L. H. (1981) ***Selltiz, Wrightsman, and Cook's Research Methods in Social Relations*** (4th edit.), New York: Holt, Rinehart and Winston, 483.

Nadler, E.D., DiSario, R., Mengert, P., & Sussman, E.D. (1990, reprinted 1992) ***A simulation study of the effects of communication delay on air traffic control,*** Report No. DOT/FAA/CT-90/6.

Natrella, M. G., (1966) ***Experimental Statistics,*** National Bureau of Standards Handbook 91, 1966.

INDEX